Vol. 68, No. 1 January 1961

Psychological Review

RICHARD L. SOLOMON, Editor
University of Pennsylvania

CONTENTS

Factorial Angles to Psychology.....J. P. GUILFORD 1

Personality Theory and the Problem of Stability and Change in Individual Behavior: An Interpersonal Approach
PAUL F. SECORD AND CARL W. BACKMAN 21

Realism of Confidence Judgements
JOE K. ADAMS AND PAULINE AUSTIN ADAMS 33

Application of Stimulus Sampling Theory to Situations Involving Social Pressure.....PATRICK SUPPES AND FRANKLIN KRASNE 46

An Analysis of GSR Conditioning
M. A. STEWART, J. A. STERN, G. WINOKUR, AND S. FREDMAN 60

Effects of Score Transformations in *Q* and *R* Factor Analysis Techniques.....DONALD M. BROVERMAN 68

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Editorial Consultants

ROBERT P. ABELSON
E. JAMES ARCHER
SOLOMON E. ASCH
ALFRED L. BALDWIN
RAYMOND B. CATTELL
WILLIAM K. ESTES
LEON FESTINGER
EUGENE GALANTER
W. R. GARNER
DAVID A. GRANT
NORMAN GUTTMAN
HERBERT C. KELMAN
JOHN I. LACEY
GARDNER LINDZEY
R. DUNCAN LUCE
RICHARD D. MANN
FREDERICK MOSTELLER
J. TREVOR PEIRCE
FRANK RESTLE
BENBOW F. RITCHIE
EDGAR H. SCHEIN
KENNETH W. SPENCE
ROGER W. SPERRY
SAUL STERNBERG
JOHN A. SWETS
JOHN W. THIBAUT
LEDYARD R. TUCKER
BENTON J. UNDERWOOD

(A large group of psychologists has kindly agreed to assist in the editing of manuscripts submitted to the *Psychological Review*. In each issue the Editor will list in the *Review* those editorial consultants who have helped him process manuscripts for that issue, so that their efforts may be gratefully acknowledged.)

The *Psychological Review* is devoted to articles of theoretical significance to any area of scientific endeavor in psychology. Except for occasional articles solicited by the Editor, manuscripts exceeding 7,500 words will not be accepted. Ordinarily manuscripts which consist primarily of original reports of research should be submitted to other journals. However, reports of original research may be included in a major theoretical article which attempts to integrate several related original studies.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Tables, footnotes, and references should appear on separate pages; all of these, as well as the text, should be typed double-spaced throughout, in all manuscripts submitted.

All manuscripts must be submitted in duplicate.

Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors should retain a copy of the manuscript, in case there is loss in the mail. Manuscripts should be addressed to the Editor, Richard L. Solomon, Psychology Department, College Hall, University of Pennsylvania, Philadelphia 4, Pennsylvania.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Streets, Lancaster, Pa.
1333 Sixteenth St. N.W., Washington 6, D. C.

$10.00 volume $2.00 issue

Second-class postage paid at Lancaster, Pa.

Send all communications, including address changes, to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must arrive by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1961 by the American Psychological Association, Inc.

VOL. 68, NO. 1 JANUARY 1961

PSYCHOLOGICAL REVIEW

FACTORIAL ANGLES TO PSYCHOLOGY[1]

J. P. GUILFORD

University of Southern California

Very rarely has it been recognized that factor analysts have anything to contribute toward the resolution of general, systematic issues in psychology. It must be admitted at the beginning that there is no unified body of psychological theory, as such, that has been developed by factor analysts. What I shall say on the subject is the opinion of one person and cannot be represented as an expression for any group.

The average factor analyst has delayed theory building because systematic thinking requires information concerning a rather large number of psychological factors before the most significant interrelationships become apparent and systems can take shape.[2] The writer had felt no particular call to consider the relation of psychological factors to behavior theory until he recently faced the need to present in book form an organized picture of personality from the point of view of factors (Guilford, 1959a). As for the systematists in psychology who are not factor analysts, probably none has regarded the method of factor analysis as a scientific tool by which psychological knowledge can be advanced very much; certainly not as a tool by which a comprehensive model of the behaving organism could be derived.

Objectives of This Paper

It is the purpose of this paper to show how factor theory and factor-analytic methods and results can provide the models and the information upon which a comprehensive theory of behavior could be based. To this end, it will be necessary to present the logical foundation for factor analysis as a scientific method and to see how we can go logically from factors to the kinds of concepts that are more familiar to psychological theory. It is much too early to say that a system of psychology has been worked out from this point of view. All I can do is to indicate some of the implications of the knowledge of factors and their interrelationships for the understanding of the behavior of individuals and some of the more general ideas that follow from these implications. The conclusions have some bearing upon many of the perennial, systematic

[1] Based upon a paper presented at the University of Florida, Gainesville, March 9, 1959, as a part of a symposium on points of view in psychology.

[2] An early attempt to develop anything like a system of psychology based upon factor analysis was that of Spearman (1923, 1927). More recently Cattell (1957) has applied factorial findings to behavior theory, culminating in a rather elaborate presentation. Neither has dealt with many of the major issues that have divided schools of thought.

issues, such as that of associationism. Thus, some of the lines of a factorial system can be drawn, and certain comparisons can be made with the major schools of psychology.

Why Factors Have Not Contributed to Psychological Theory

One of the reasons the findings of factor analysis have not contributed more to psychological theory is that many analyses have not been basic studies of the properties of human nature, either by intention or by way of outcome. One practical use of factor analysis is the reduction of correlational data, in order to transform them into simpler form for inspection. The chances of revealing anything of fundamental psychological interest under these circumstances are rather remote.

Even when the objective is to discover something of general interest regarding human nature, the outcomes are often disappointing. Many factorial studies of this type are poorly planned because they are forced to be simply exploratory. In this instance, there is not enough prior information as a basis for setting up fruitful hypotheses. Even a single, prior exploratory analysis in the same area of behavior would be of considerable help. Without good hypotheses there is limitation as to the wise choice of experimental variables. The choice of variables to be analyzed together is all important for success. We cannot go into the technical reasons for this statement here, but the well-worn cliché "You get out of a factor analysis what you put into it" does apply, if it is revised to say: "You *cannot* get out of an analysis what you do *not* put into it."

Another reason for some failures of factor-analytical studies is the lack of good experimental design. More will be said about this point later, but I should like to add here that the fact that we have an elaborate computational procedure for treating data does not relieve us of the necessity of observing the ordinary requirements of experimental logic and experimental controls.

But results from the better factor analyses are also commonly rejected as contributions to general psychological knowledge. The reason is probably that the method has been most often used for the study of individual differences. The fact that these studies have often been inspired more by the interests of vocational psychologists than by interest in basic facts of behavior is relatively coincidental. The basic data with which the factor analyst starts are measurements of numerous individuals in a large number of experimental variables. In other words, the emphasis is upon individual differences. Factors, when interpreted psychologically, are also conceived as ways in which individuals differ from one another. Psychological theorists, on the other hand, are primarily interested in how individuals are *alike* and only incidentally in how they *differ*. It is actually very easy to bridge the gap, but there has been little effort to do so.

In this connection, the factorist could well charge the general theorist with being negligent for not bringing individual differences into the picture when he constructs a system of psychology. No well-known system, with the exception of Tolman's purposive behaviorism, has anything of significance to say about individual differences. Individual differences are recognized as empirical facts, but they have been generally treated as nuisances by the experimental psychologist, and as phenomena of little or no interest by the theorist. The fact

that reasonably satisfying systems could be built ignoring individual differences is an important reason for not bringing them into the picture.

One consequence of this general state of affairs is that the theorist who is concerned with personality often finds such systems inadequate for his purposes. Consequently he builds his own theory. But even such a theory often recognizes only implicitly the need for a significant emphasis upon individual differences. It rarely does justice to this aspect of the meaning of personality. Such a theorist recognizes that each person is unique but he does not sufficiently appreciate the fact that it is individual differences that make him unique.

SOME FACTOR MODELS FOR PERSONALITY

Factor theory has provided the only rigorous models of any consequence that have been proposed to make intelligible the many facets of the phenomena of individual differences. I shall mention three of those models because they are especially relevant to this paper.[3] A multidimensional model is used to represent personality in general. A hierarchial model represents interrelationships among traits within typical individuals. A matrix model represents systematic relationships among factors within a certain domain of behavior.

A Dimensional Model

The most common, basic model is a multidimensional Euclidian space, each dimension of which represents a unique trait. A unique trait can be discovered as a common factor by factor analysis. To the extent that an individual's personality can be accounted for in terms of a limited number of common factors, each person can be represented as a point in n-dimensional space.

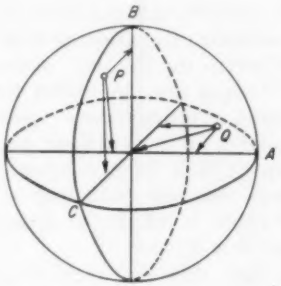


FIG. 1. Example of a dimensional model involving three unique traits, or common factors, A, B, and C, showing characteristic point positions of two individuals, P and Q, with respect to the three personality traits.

Figure 1 illustrates this kind of model, showing only three dimensions, for obvious reasons. Each axis is a linear dimension representing a unique trait along which individuals have characteristic trait positions. Let us say that axes A, B, and C represent the disparate traits of gregariousness, meticulousness, and emotional stability, with their positive extreme qualities at the labeled ends of the axes. So far as these three traits go, Individual P is described quantitatively by his projections on the three axes. His three projections define a point, which is his characteristic position in this three-dimensional space. The point for Person P tells us that he is a bit below average in gregariousness and that he is very meticulous and also moderately strong in emotional stability, the average for each trait being at the origin.

Person Q has another combination of three projections, defining another point location in the same space. In n-dimensional space each person has n different projections defining a point in that n-dimensional space. The fact

[3] Still other models have been presented by Guttman (1954) and Lazarsfeld (1954).

that he probably changes his trait positions somewhat from time to time need not concern us in this discussion. Such changes are the reason for the expression "characteristic position" used above. If it is difficult to think of a space of a large number of dimensions, one can think instead of a profile chart, in which the dimensions are laid side by side.

The fact that the axes shown in Figure 1 are orthogonal (at right angles) is incidental. Actually, factorists generally agree that the best-representing axes more often than not depart from right angles. They do not agree as to how the angular separations shall be determined empirically.

A Hierarchical Model

Although the dimensional model just seen provides for the description of individuals, it is better as a representation of variables within populations than as a picture of personality structures within individuals. There is also the circumstance that factor traits are of different degrees of generality, in the sense that some are related to wide ranges of behavior and some to narrower ranges.

A brief example will illustrate. Let us say that Student Z declined an opportunity to cheat while taking a certain examination. This is a specific action; there is no necessary implication concerning Z's traits from this one observable event. Over a run of 20 consecutive examinations, card games, and purchases Z also passed up 18 opportunities to cheat. We may infer that over this range of behavior Z has a rather high position on a trait of resistance to cheating. So far as we know from these observations, the trait is not a very broad or general one. Further observations show that Z's conduct is similar in situations that offer opportunities for deceiving and stealing. From this information we may conclude that he is strong on a more general trait that we may call "honesty." But Z has an even more general tendency toward ethical conduct of other varieties, which justifies our placing him above the average on a still broader trait of "strength of character." Furthermore, he exhibits an even more general restraint and self-control that goes beyond behavior that has ethical implications, suggesting a still broader disposition or syndrome.

Because of the appearance of traits at different levels of generality, a few writers feel the need of a second model to supplement the first (Burt, 1949; Eysenck, 1951; Guilford, 1959a; Vernon, 1950). This is a hierarchical model, which serves to relate the factorial dimensions logically to one another and to a single personality. Figure 2 is an example of a segment of such a model. In the illustration of traits just given, the particular acts of resistance, in whatever area of behavior—cheating, deceiving, or stealing—are at the specific-action level. Traits of cheating, or of deceiving, or of stealing may be regarded as being at the "hexis" level. A hexis is about the same as a habit but the term "hexis" is preferred because it is not so committed to the idea that traits are determined by learning only. A trait of honesty, for which there is factor-analytical evidence (Guilford, 1959a), may be considered to be at the primary-trait level. Strength of character and general self-restraint would be considered syndrome types at two still higher levels.

Factor analysis can be applied at various levels in the hierarchy of traits. From information concerning intercorrelations of measured specific ac-

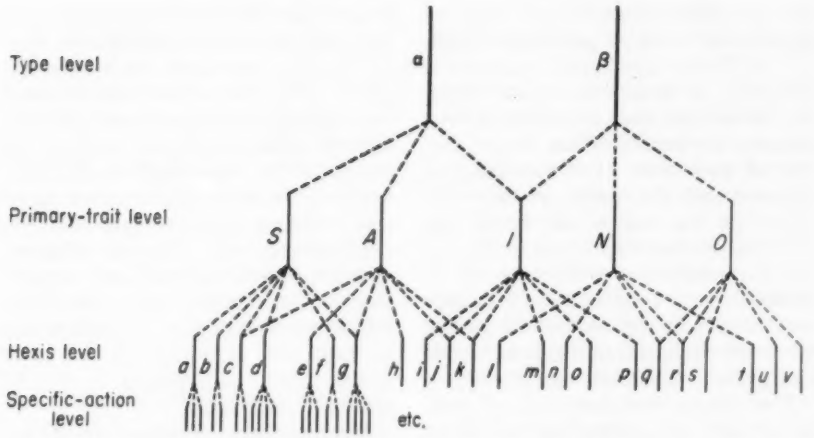


FIG. 2. Segment of a hierarchical model representing an area of personality, with three levels of traits of differing generality and the specific-action level.

tions we can derive conclusions regarding the existence of unique traits at the hexis level. From information concerning intercorrelations of trait positions of individuals for traits at the hexis level, we can derive conclusions regarding the existence of unique traits at the primary-trait level; and so on, for higher levels. Probably most of the psychologically meaningful factors thus far discovered have been at the primary-trait level.

Many of the apparent disagreements in results from different analysts should be attributed to the fact that they are analyzing at different levels in the hierarchy. The kind of factors one obtains depends upon the level or levels at which he obtains his information concerning individuals. In some analyses, factors of different levels are undoubtedly confused. From the standpoint of clarity of results the best general strategy would have been to start with information from the lower levels and work upward. Before we can collect clear evidence regarding status of individuals on unique traits at a certain level, we need to know what are the unique traits at that level.

The Matrix Type of Model

The third model is of recent origin and comes about from attempts to discern logical interrelationships among the known factors. Since the factors concerned are probably at the primary-trait level, the resulting models apply within that level. The results at the primary-trait level reveal organizations of the factors that suggest relations to higher levels. Hence the general outcome of this kind of model building may serve as guiding hypotheses concerning the completion of the hierarchical picture of personality. The best evidence concerning interrelationships of the levels will be empirical information regarding intercorrelations of factors. As yet we have no adequate experimental procedures for estimating these correlations.

Attempts have been made to classify the known factors in certain areas of personality, with some success in a few of them, including the psychomotor abilities, the intellectual abilities,

the temperament traits, and some of the simpler traits of pathology (Guilford, 1959a). The typical outcome is a matrix of factors that places them in columns and rows according to their common properties. That is, the factors of each row have something in common and the factors of each column. In the case of the intellectual abilities, of which about 55 are known at the present time, a third dimension is necessary. They have been organized in a solid figure, known as the structure of intellect, which is shown in Figure 3 (see Guilford, 1959b).

For our present purposes, we need to note only the general features of the model of the structure of intellect. First, the factors are classified according to the kind of operation performed: *cognition, memory, divergent production, convergent production,* and *evaluation.* Second, the factors are classified according to the kind of content or material on which the operations are performed—*figural, symbolic, semantic,* and *behavioral.* Of the four content categories we might say that the first represents concrete intelligence, the next two represent two varieties of abstract intelligence, and the fourth represents social intelligence. The latter has been included on a purely theoretical basis; no factors of social intelligence are as yet known. The third basis of classification is in terms of the kinds of products resulting from the operations on the contents—*units, classes, relations, systems, transformations,* and *implications.* Theoretically, there should be a different primary intellectual ability in each cell of Figure 3, in other words, about 120 instead of the 55 presently recognized.

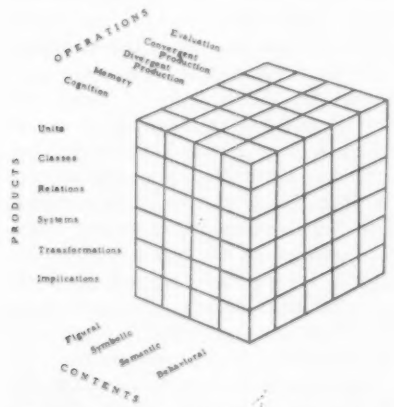


FIG. 3. Cubical model of the structure of intellect, representing categories of primary abilities with respect to three modes of variation.

We shall have occasion to refer back to the structure of intellect, for I shall draw heavily upon the principles that it demonstrates and upon the implications drawn from it in connection with discussion of general psychological theory. This is so, not only because intellect looms large among the things to be accounted for but also because our knowledge from the factorial approach is extensive in this area of personality.

FACTOR ANALYSIS AS A TOOL IN SCIENCE

Factor analysis is a potent instrument for extracting information from data, but it has no magical power to reveal anything regarding information that is not inherent in the data. The scientist who uses the method for the discovery of psychological information should start out by asking certain questions before he collects his data. In other words, a factor-analytic study must be carefully planned, with clearly stated hypotheses to be tested, if one is to use the method effectively.

The more experimental features the investigator can bring into his planning the better. This practice reduces ambiguity in the interpretation of results. In evaluating the scientific use

of factor analysis, therefore, we need to be clear as to the kinds of hypotheses that can be set up, the ways in which those hypotheses are tested, the kinds of variations of conditions introduced, and the kinds of controls exerted. Many psychologists are probably unaware that these experimental features are possible in factor-analytical investigations, perhaps because in publications regarding the method so much attention is given to computational steps and so little to experimental uses of it.

A Typical Factorial Study

Let us illustrate an experimental design for a factorial investigation by reference to a study of intellectual abilities. Suppose we have selected for study the area of problem solving. Suppose that we had no prior information regarding primary abilities involved in problem solving. Our preliminary thinking about problem solving would first probably suggest two alternative hypotheses. One would be that there is a single, universal ability to solve problems and the alternative would be that there are a number of distinctly different abilities involved. If these were the only hypotheses that we wished to test in our first analysis, we would simply select a list of tests, each of which presents problems of a different type. The types would be varied considerably, covering a great range. The analysis should give us the answer as to whether individual differences in success in solving these many different kinds of problems can be accounted for on the basis of one common factor or whether we are led by the results to reject this idea.

Let us say that the results indicate the acceptance of the multiple-factor hypothesis to account for problem-solving ability. The next study would normally be directed toward answering the question of how many factors are needed and the question of the properties of each factor. The hypotheses set up in the second investigation would probably involve a prediction that a certain number of factors is needed and that each expected factor has certain properties. The investigator would be aided considerably in this hypothesis formation from the way in which the tests were classified among the factors in the first analysis. But there might be ambiguities regarding the properties of each factor. Hypothesized factor Q might be expected to have properties a, d, and f, or it might have properties c, d, and f, or even properties a, c, and g. From the hypothesized properties arise ideas as to the kinds of tests that will be needed to measure the factor. Experimental tests are accordingly constructed, differing systematically along the lines of the supposed properties, with the hope that when a hypothesized factor does come out in the analysis, tests with one kind of specification will be found related to it and tests with the other specifications will not. In this manner we hope to obtain clarification of the nature of a factor.

Statistical Tests and Factor Analysis

At this point the rigorously inclined investigator is quick to remind us that there are no adequate statistical tests that can be applied, either to help us decide how many common factors obtain in the particular analysis or whether factor loadings are significantly different from zero or from one another. There is no use denying this state of affairs. It is also true that there is some degree of looseness connected with rotational procedures, which also contributes to the uncertainty of conclusions. Conclusions must therefore be made on a permis-

sive rather than a compulsory basis, in spite of the fact that an accepted rotational solution appears to be compulsory.[4] To be completely objective, we can only say that yes, we may tolerate the idea that there is a unique psychological trait having such and such properties, in line with preliminary hypotheses, or we may say that no, we cannot reasonably do so. In the latter event we may conclude that the factor has other properties not suspected before and a new analysis is called for.

The reaction of some psychologists to this state of affairs may be to reject the results from factor analysis completely. To such a person it can be said that much more important to a scientific psychologist than statistical tests are psychological ideas. Sciences generally, including psychology, originally developed and went a long way without the aid of statistical tests. The lack of statistical tests is not fatal, but the lack of ideas would be. This is not to say that we should not wish to have both if we can. But there should be no prohibition against applying a method that seems fruitful, while waiting for statistical tests to be developed.[5] I suggest that we let the results of factor analysis speak for themselves and that they be judged on the basis of how much they contribute to psychological understanding, prediction, and control.

[4] A number of proposals have been made for rotation according to rigorously defined criteria by analytical methods. Although some of them are very attractive in principle, they usually fail to achieve fully satisfactory solutions from an intuitive point of view. The difficulty is that the selection of experimental variables must be nearly ideal in order to achieve the "right" solution from an interpretive standpoint.

[5] Some progress is being made with regard to statistical tests of certain aspects of factor-analytical results (see Maxwell, 1959).

Experimental Controls in Factor Analysis

To return to other experimental features of a factorial study, what conditions are systematically varied? The most conspicuous kind of variation is in the types of tests, or other experimental variables, used. For the most part these variations are qualitative. From test to test we find variations in kind of material—figures, pictured objects, letters, numbers, and words being the most common varieties in printed form. We find variations in item format—multiple-choice, matching, completion, and their derivatives. We find variations in instruction as to what is to be done and how it is to be done. In some of our more recent analyses we have introduced a few *quantitative* variations, such as the number of restrictions of a certain kind, the number of responses called for per item, and difficulty level of tests with the same kind of items. In connection with the quantitative variations we sometimes predicted that the factor loading for a certain factor would rise systematically and that for another factor would fall systematically, or that there would be an intermediate optimal level for the variation concerned. A great deal has to be known about the nature of a factor before this much refinement in experimental variation can be effectively introduced.

Certain conditions are at least partially controlled, or should be. The selection of population of individuals is important. In the study of intellectual abilities, there should be relative homogeneity in age, education, sex, and general intellectual level. Ideally, one should like to equate individuals on all other factors than those under investigation, but this would be prohibitive in terms of effort, and fortu-

nately it is not essential. We can tolerate incomplete controls, as must be true of any experiment, allowing the effects to go into the error components. Factor analysis has a way of segregating error variance from the variances in which we are interested, not perfectly, but within practical limits. So long as the non-error variance is of sufficient potency, we can obtain a fair picture of the common factors.

FROM FACTORS TO FUNCTIONS

Our next need in this discussion is to establish a logical bridge between factors and general psychological theory. Present psychological theory has arisen mainly in connection with the traditional experimental approach, hence in constructing the bridge we should profit by giving some attention to a comparison of that approach with the factor-analytical approach, in certain respects.

Individual Differences vs. Functioning Individuals

The main operational difference is that the traditional experimentalist focuses his interest upon stimulus-response relationships. The stimulus-response model is basic to his thinking and to his planning of investigations. The factorist, on the other hand, directs his attention primarily upon responses and the concomitances among responses. From the intercorrelation of response values the factorist looks for signs of traits, not for stimulus-response relations. Traits are properties of individuals. To the student of personality, their determination of behavior is just as real a phenomenon as the determination by stimuli. Some experimentalists are now recognizing this general principle, as indicated in studies of the relation of characteristic anxiety levels to various measured effects in behavior. In such an experiment, one independent variable is not a stimulus variable or even a transitory personal variable such as a mental set. It is individual differences on an enduring attribute, in other words a trait.

The key to the bridge, therefore, is the concept of "trait." In the context of personality theory, a trait is any relatively enduring way in which one person differs from others. On a scalable trait, which can be represented by a straight line, each person has a characteristic position. If individuals have different positions on a common scale, the scale represents some quality or property that each person possesses to some degree, in common with other persons. If the quality is a unique one, such as may be discovered by factor analysis, it is some significant component of the individual's constitution.

There is nothing in this line of reasoning to force us to any conclusion as to the fundamental nature of this component. It might be a unique skill, a unique motive, or a unique attitude of some kind. It might be dependent upon some particular organic structure or upon some combination of structures functioning together. There is nothing in ordinary factor-analytic results to tell us about the origins of a factor or to give us sufficient basis for its automatic classification.

Traits and Functions

If we consider the primary abilities represented in the structure of intellect (Figure 3), however, and ask ourselves what kinds of components they are, we obtain a little help. Examination of the list of intellectual abilities tells us that represented in one particular cell is an ability to cognize (discover, know, or rediscover) semantic systems and that this is distinctly dif-

ferent from an ability to evaluate the logical soundness of conclusions involving relations between symbols. If each person has a characteristic level of ability to perform in each of these two respects, he certainly is performing in these two ways. In other words, we may say that he has *functions* of these two types. The primary intellectual abilities may therefore be conceived as ways of functioning within individuals as well as ways in which individuals differ from one another. Thus, by the study of how persons differ we also discover how they are alike.

Traits other than abilities are not so readily conceived as functions, but it is possible to see that some of them are the *directions* that the functioning individual takes (in the case of motivational traits) and others are connected with the *manner* in which the functioning proceeds (in the case of temperamental traits). As in the case of abilities, other primary traits should give us ideas about the functioning of the individual as a going concern. Information regarding the factors and their properties should therefore help us to build up a picture of the behaving organism, and hence contribute to psychological theory.

SOME IMPLICATIONS OF FACTORS FOR PSYCHOLOGICAL THEORY

It has just been suggested that from the psychologically meaningful factors that we know, and from their interrelationships, we can make progress toward the general description of behavior of individuals and account for the determination of behavior by internal sources. The primary traits that are recognized as needs, interests, and attitudes, provide potentially useful concepts for understanding the varieties of general motives operating within the individual. Forty or more years ago it was common for psychologists to draw up lists of instincts. In later years this kind of exercise has been replaced by a somewhat less common effort to draw up a list of fundamental human motives. McDougall, for example, took this kind of objective very seriously. If the point of view adopted in psychology calls for such an analysis of motivation into component urges, propensities, needs —call them what you will—factor analysis offers a very natural empirical method for arriving at such a goal.

Information, Operations, and Products

From the intellectual factors as organized in the structure of intellect, some of the most significant implications for psychological theory may be seen at this time. The first general principle is that we should conceive of a human being as an instrument or agent whose psychological purpose is to deal with information. In saying this, it is necessary to interpret information in a very broad sense. For our present purposes, "information" may be defined as anything that the individual discriminates. He need only show that he discriminates or can do so by means of differential reactions of a sufficiently consistent character.

This principle regarding information and the definition of information do not follow directly from the structure of intellect, but they are strongly suggested by it. The structure of intellect does indicate that the individual cognizes, which means that he discovers or rediscovers or recognizes information. The information may be in figural form, which is a form that can be perceived through the senses. It may be in symbolic form, which may involve letters, digits, or other conventional signs. It may be in the form of meanings, ideas, or concepts, or what I have called semantic content.

If the behavioral column is appropriately included in the structure of intellect, the information may be in terms of such things as the intentions, feelings, and attitudes of other individuals and of ourselves.

The individual not only cognizes all these kinds of things, he also retains information for future use. According to the structure of intellect, for every cognitive ability there should be a corresponding ability to retain information. Eight such memory factors are known. The individual also generates new information from given information, in accordance with the two categories of productive operations—divergent and convergent. Finally, the individual evaluates information, whether it is merely cognized or whether remembered or produced from other information. He decides whether the information is correct, adequate, suitable, and so on. A negative decision provides the occasion for new activity involving the same kinds of operations, with the individual probably arriving at new products.

The kinds of products achieved are as indicated by the six categories of factors known at present. Units are relatively segregated items of information and they are sometimes articulations or integrations of parts. The latter characteristic is best seen in the symbolic column of factors, in connection with which units may be in the form of syllables, words, or number combinations. By virtue of common properties among the units, classes can be formed. Units also enter into relations with one another, for example, relations such as larger than, attached to, or inside of. One might think that systems are merely complex sets of relations, but analysis shows that abilities needed to cope with them are different from the abilities needed to cope with relations. Transformations include changes of various kinds, for example, changes in spatial order or arrangement, in letter combinations, and in the uses or functions of objects. Implications are in the form of extrapolations from given information, such as extensions of the information in space or in time. Thus, the six kinds of products seem to pertain to varieties of knowledge when knowledge is reduced to elemental terms.

An Informational Theory of Learning

A psychological system should provide the principles by which the facts of various areas may be ordered and comprehended. As an example of how the factorial view would account for an area of behavior, let us consider the subject of learning. Each point of view should have something to say regarding this most important subject. What kind of theory of learning would be consistent with the known outcomes of factor analysis or would be implied by this information?

Taking our cue from the structure of intellect, we find that the theory of learning suggested is in the general category of cognitive theories. Furthermore, from the list of cognitive factors we see the classes into which such cognitions fall. Learning, then, is essentially discovery. The particular discovery may be achieved completely and suddenly, an instance of learning that has been called "insight," or it may be achieved in a series of approximations. The discovery may be in any of the six forms recognized as products in the structure of intellect. The achieved cognition may be an acquaintance with a new unit, the formation of a new class, the formation of a new relationship or a new system, the awareness of a transformation, or the extension to new implications.

It is possible to give an associative interpretation to some of these events

but not all. Perhaps the most obvious type of product that may be interpreted as an association is that of relation. But when we recognize that two things bear some relationship to one another, it is not just that when one comes we think of the other. The relation itself is something additional. It may be a difference along some continuum, as when we say that X is taller than Y, P is older than Q, or J is softer than K. The relation might be in terms of spatial or temporal position. Or it might be in terms of opposition, friendship, status, or of whole and part. One particular relation may be of very general importance, namely, the relation to the effect that "this belongs with that" merely because the two things occur in temporal or spatial proximity. There may be no other basis, for example, in terms of properties of the two things, for justifying the formation of a relationship.

It is possible to apply the concept of association in connection with the cognition of classes. Classes can be formed because units have something in common. We could say that Aristotle's concept of association by similarity applies. But being aware that a class exists and knowing a class concept would seem to go well beyond a mere association by virtue of the similarity of two things.

The more modern interpretation of class recognition brings such an event under the concept of "equivalence of stimuli." But there seems to be a genuine psychological difference between failing to discriminate stimuli, on the one hand, and cognizing that discriminated stimuli have properties in common and therefore belong in the same class, on the other. We could think of all such events, of course, as belonging on a continuum of lower to higher levels of cognition of classes. The implication of this discussion is that class recognition, however primitive or sophisticated, is still a unique kind of psychological event.

A third possible place for applying the associative principle is in connection with the product of implication. One of the natural conditions for developing a cognition of implication would seem to be temporal sequences of stimulations. It seems likely that what some learning theorists refer to as "expectations" come within the category of implications.

In connection with the other three kinds of products—units, systems, and transformations—the association principle seems *entirely* inappropriate. Wundt and others failed in their attempts to account for perceived objects in terms of association. An important natural consequence was the development of Gestalt psychology. Units, systems, and transformations all pertain to things that have totality properties that defy associative interpretations. As a matter of fact, it can be maintained that even in the products of relations, classes, and implications, totality properties are also involved. Two things in relation constitute a kind of unitary structure, as Spearman pointed out. Classes are conceived groups of units and to each group a single class concept is applied. The emergence of a class idea is hardly to be accounted for as the formation of an association. In connection with implications we often have *continuities* of various kinds. Since it is possible to conceive of totality properties for all six kinds of products, it is small wonder that the Gestalt psychologists felt considerable confidence in applying their principles to all kinds of mental functioning, with complete discard of the principle of association.

As a particular example of human learning, let us take the case of memorizing a series of nonsense syllables,

so popular traditionally in the laboratory. From the factorial point of view, each syllable is a symbolic *unit*, and the learner becomes more and more familiar with it as exposures are repeated. In learning to associate each syllable with the one following in the list, the learner, we may say, is seeking for and perhaps discovering a *relation* (or relations) where none existed before; a relationship may be formed involving two units of information. Two syllables offer very little in the way of properties that can be utilized for this purpose, as compared with words. But each pair of syllables does offer some possibility of the formation of some kind of relationship.

Another aspect may be that the learner is acquiring new expectations or *implications*. This may be indicated by the fact that forward connections are stronger than backward associations. We may also hypothesize that the learner is developing a new *system* of units. At least two primary abilities that we know indicate that order is an important kind of system. Some indication that a system is being formed can be seen in what are commonly called remote forward associations and backward associations. The beginning and end units of the series of syllables seem to serve as anchoring points for the whole system and hence have much to do with the relative ease of learning units in different positions in the series.

This line of reasoning suggests that in memorizing a list of syllables several of the primary abilities come into play, including the cognition of symbolic units, of symbolic relations, of symbolic systems, and of symbolic implications. Since the testing of progress in learning involves lapse of time, there is probably also involvement of the corresponding memory factors. If the series were composed of figures or of words, the corresponding cognitive and memory factors in the figural and semantic columns, respectively, should be involved. Still other intellectual resources may come into play, depending upon the strategy of the learner. The hypotheses concerning the roles of various common factors in a particular learning task can be tested by factor analysis of scores obtained at different stages of learning along with sufficient marker tests for the factors probably involved. Fleishman and Hempel (1954, 1955) have already conducted studies of this kind.

Reinforcement

Most theories of learning have some provision to account for the fact that practiced changes in behavior have some degree of permanence, in other words, some concept to account for what is called "fixation." The concepts usually invoked for this purpose are some form of a law of effect or of reinforcement. In an informational theory of learning based upon the structure of intellect, the key to fixation is to be found in the category of evaluation. In this connection, feedback information is very important.

Thorndike seemed to favor the interpretation that effect is in the form of a confirming reaction, a conception that seems close to that of evaluation. The idea of confirming reaction is not comprehensive, however, because it does not give significance to *failure* of confirmation. In stressing the positive aspect of evaluation, Thorndike was reflecting the experimental findings, as he thought, that only positive effect seemed to change the strength of learned connections. He did recognize that a failure leads to new learning activity, which is equivalent to saying that an evaluation of failure is the occasion for new learning effort.

The conception of an evaluative

aspect to learning is broad enough to embrace most current interpretations of effect. When primary reinforcement is equated to drive reduction, we can say that the organism is using feedback information regarding his internal condition. Secondary reinforcement may be said to involve *implications* cognized by the organism. Lest these statements raise objections that there cannot be unconscious cognitions and evaluations, as a sort of contradiction of terms, let us refer back to the definition of information that was given earlier—information is that which the organism discriminates. The conscious-unconscious distinction is therefore irrelevant.

The evaluative view of effect just expressed suggests the extension of the application of the concept of information to the phenomena of motivation and emotion generally. This extension leads to some new ways of looking at old problems. Both motives and emotions, including feelings, can be regarded as information concerning the internal conditions of the organism; concerning its needs and its states of well being or ill being.

In this connection we need to recognize another property of information. When motives become at all strong, there is a sense of urgency about them and about the accompanying emotional aspects. This suggests the need for the concept of "loading." We can say that information carries different degrees of loading or weight. Feelings of pleasure and of displeasure are in themselves information that is more or less loaded, positively or negatively. There is increasing evidence that in themselves these feelings can be reinforcing (Miller, 1955). They are not infallible: information and evaluation may sometimes go against them because they are outweighed by other information, but they nevertheless can be regarded as kinds of evaluative information. In this connection, it may be of more than incidental significance that in the work of Osgood and others on the dimensions of meaning, the pleasant–unpleasant or evaluative dimension seems to be the strongest and most universal (Osgood, 1957).[6]

Another thought on "loading" would be that the basis for it is to be found in the behavioral column of the structure of intellect. Information regarding our own internal conditions and our relation to external situations is evaluated as being vital to some degree. The degree of recognized vitalness determines the loading or *is* the loading. We are thus brought to the conclusion that loading is a matter of evaluation and that it may apply not only to behavioral products but quite generally.

Transfer

The general problem of transfer of learning takes on a different complexion in view of our information concerning factors. G. A. Ferguson (1954) has proposed the unique theory that factors of ability are produced by the fact that transfer of learning takes place along the lines of the common factors. It has long been recognized that psychological similarity of tasks is a key to the direction and extent of transfer. Similarity of tasks depends upon common kinds of content, operations, and products, or of combinations of these features. We interpret common factors in terms of the common properties of the tests that indicate them.

[6] It is also significant that J. P. Seward (1958) has recently assembled considerable evidence in support of the cognitive interpretation of learning and also that there is evidence pointing toward need satisfaction that arises merely from acquiring new cognitions.

Ferguson's theory does not necessarily mean that all the effects of learning are general. It is possible to deduce from factor theory, as applied to abilities, the principle that each instance of learning has both general and specific components. Fleishman and Hempel (1954, 1955) have provided evidence to support this principle from more than one investigation. Each time, they found a specific component, whose relative importance for individual differences systematically increased with practice, in addition to the common-factor components each of which tended to undergo some systematic changes in importance. The common factors were identifiable with those found in testing apart from the learning experiment.

Although the Fleishman and Hempel experiments give us no information regarding progressive improvements in any of the common factors, they demonstrate that common factors do play roles in learning and they find no evidence of the existence of a unique, general learning-ability factor. These and other factorial learning studies should nail in its coffin for all time the notion that there is a single, general learning ability. Abilities to learn are apparently at least as numerous as there are intellectual factors. The writer often feels that psychologists would do well to banish the term "learning" entirely and instead talk about changes in behavior, their conditions and consequences.

Factors in Problem Solving

The same fate is in store for the notion that there is a single problem-solving ability. As I have pointed out elsewhere (Guilford, 1954), the variety of problems is enormous. Each type of problem, like each type of task, is presented in the form of a certain kind of content, it calls for its own kinds of operations, and certain kinds of products are needed in order to achieve a solution. A problem exists for an individual when to cope with a given situation he has no adequate, previously learned and remembered device ready to deal with the situation. From this definition we see that problem solving is as broad as behavior itself. It is likely, therefore, to tax any of the resources of intellect that may be pertinent.

Each type of problem should draw upon its own pattern of resources in the average person in a certain population. The importance of each factor for individual differences in problem-solving scores can be indicated by a factor analysis of those scores along with scores from appropriate marker tests of the factors. In our Aptitudes Project at the University of Southern California,[7] we have just completed this type of study for one type of problem. Five factors appeared to be of some importance: *sensitivity to problems, verbal comprehension, ideational fluency, originality,* and *conceptual foresight.* The conditions, as represented by the problem tasks and the intellectual status of the subjects and their strategies, were evidently right for emphasizing these five factors.

Motor and Sensorimotor Aspects of Learning

Many readers will recognize that the account of learning is limited in a number of ways, because only the intellectual factors have been considered, and it needs certain supplementations. The theory of learning presented is not considered a complete

[7] Under Contract Nonr-228(20) with the Office of Naval Research, monitored by the Personnel and Training Branch. This particular study was under the direct supervision of Philip R. Merrifield (Merrifield, Guilford, Christensen, & Frick, 1960).

one by any means. There are questions of how sequences of cognitions and actions are developed and how complex motor patterns are organized as a consequence of activity. Two individuals who have the same cognition may not do the same thing or do it in the same manner. Adequate motor responses are not always available on demand and new ones must often be developed.

From a cognitive point of view, sensorimotor learning can be conceived as the formation of cognition-response sequences instead of stimulus-response sequences. If we knew the individual's cognitions, we should be better able to predict his responses than we can from knowledge of the stimulating conditions. He responds overtly to his cognitions rather than to the experimenter's "real" world. For a long time the Gestalters have been attempting to convince us of the validity of this principle. If the learner's cognitions nearly coincide with the experimenter's, of course, the experimenter's knowledge of stimulating conditions is sufficient for practical purposes. But where there are notable differences between the two views, it should pay to attempt to learn what we can regarding the learner's cognitions. The learner's evaluations would be other important sources of information regarding what he will do and with what energy.

The other problem, that of development of motor patterns, can possibly be approached profitably by giving attention to kinesthetic cognitions as a subject for study. The learner has a flood of information from himself as well as from his external environment, information of which the experimenter is unaware. A systematic factorial exploration of the individual's use of his kinesthetic content should be fruitful. It should tell us whether information from this general source also comes in the form of the six kinds of products and what sorts of things these products are in the kinesthetic domain. The individual's information concerning his responding equipment, its potentialities and limitations, from this source, should have an important bearing upon what he does with that equipment.

RELATIONS TO OTHER POINTS OF VIEW

Some Major Issues

Before comparing the point of view presented here with particular schools of thought, let us consider its position on a few of the major issues that divide those schools. One of the issues is sometimes called "mentalism," and it is concerned with the question whether conscious experience is accepted as a phenomenon and whether it plays a significant role in a system. My impression is that a system developed from factorial information and thinking need not take sides on this issue; the issue is not so very relevant. Actually, a completely objective system could be worked out, as one can see from the way in which the application to phenomena of learning and problem solving needed no reference to conscious experience or to concepts necessarily implying it. But if conscious experience is admitted as a necessary problem for psychologists, the factorial approach has much to offer in determining the dimensions of that experience and hence in finding descriptive concepts for it.

The assertion that the factorial view can generate an objective system needs additional defense, for the terms "cognition" and "thinking," particularly, arose in the context of mentalistic psychology. In spite of their origin, these terms, and others used in connection

with the structure of intellect, can be given entirely objective interpretations, for the reason that they have observable referents. The referents are the test materials used in factorial investigations. The factors, and in turn the factor categories, are interpreted in terms of the kinds of test content, the kinds of operations demanded by the instructions and the items, and the kinds of responses needed to make good scores.

In one sense, factor analysis is analogous to the atomic physicist's cloud-chamber procedures. The physicist draws a conclusion regarding things that he cannot observe, from observed events remotely connected with them. From such scattered sources of observation he builds up a model of the atom and of its nucleus. From test scores and the ways in which they are interrelated, the factor analyst constructs conceptions of human nature. The one seems as objective as the other.

The issue of associationism has already been touched upon in discussing the nature of learning and it will be recalled that the outcome was to reject association as a principle that would account for the products of information. There are important areas of intellectual functioning in which development of understanding has been seriously handicapped by our staking so much on one principle. Because the principle has served in so many places, in spite of objections and necessary qualifications and omissions, overreliance on the principle has continued. When the further restriction of recognizing only stimulus-response associations is imposed, psychology is further limited. This brings us to another issue.

A third general issue may be called "centralism vs. peripheralism." The extreme peripheralist not only restricts his observations to stimuli and responses, but also chooses to limit himself to descriptive concepts that pertain only to those events. The centralist, on the other hand, although recognizing that we are limited to observations of stimuli and responses, does not hesitate to attempt to infer something regarding processes that take place between those events. Hull and others have seen the need for inferring a few intervening events or conditions (called intervening variables) but they attempt to get along with a bare minimum.

It is my conviction that a psychology that restricts itself entirely to the task of correlating stimuli and responses and that avoids attempts to build up a picture of what goes on between them will fall far short of complete success. The most significant things happen between stimuli and responses. It is these things that make known stimulus-response correlations most significant. In predicting behavior we need to know about internal conditions, including traits, as well as about the external situation. The same is true concerning the control of behavior. The factor-analytic approach is aimed at a conceptualization of psychological properties of a central nature. Whether it can supply all the kinds of concepts that are needed for this purpose remains to be seen.

Comparisons with Some Major Schools

From the preceding discussion it should be clear that a psychology based upon factor analysis agrees with behaviorism in achieving an objective science but it finds the behaviorist's exclusive dependence upon the association principle to be a serious deficiency, and behaviorism's restriction to peripheralism also seriously limiting.

A factor psychology finds in the Gestalt point of view quite a number

of congenial features. As pointed out earlier, it finds the Gestalt emphasis upon totalities to be useful as a replacement for associationism. It finds, however, that Gestalt psychology failed to carry through its own program. In part this might have been due to its rapt admiration for totalities and to its aversion to analysis. Gestalt psychologists, by further elaboration in their thinking and by experimental approaches might have achieved a set of concepts such as those that describe the structure of intellect, but perhaps for lack of a method like factor analysis they would not have succeeded. Factor theory, with its additive equations, is no doubt anathema to Gestalt psychologists if they pay any attention at all to that theory. In spite of the summative equations of factor theory, factor analysis has not destroyed totalities but has discovered what many of them are like, for example, the products of information mentioned earlier.

In terms of general views and ways of thinking, the kind of factorial psychology of which I have been speaking comes closest to the traditional point of view of functionalism. The easy transition from factors to functions is the simplest indication of this similarity in type of fundamental thinking. Functional psychology never became very systematic, but through factor-analytic findings it could do so. One difference in the two views is that the functionalist has often emphasized Darwinian concepts as explanatory principles. This kind of explanation is not essential to a factorial point of view.

And what of the "dynamic" psychologies? Among this group we generally recognize the psychoanalysts, hormic psychologists, and the Lewinian psychologists. The strong intellectualistic bias of the discussion in this paper, which took its start from the structure of intellect, should not mean that factor analysis has nothing to offer regarding dynamic aspects of behavior. I have skirted this aspect, suggesting that motivation and emotion can be related to the domain of intellect in connection with evaluation. Another important point of contact would be the category of hypothesized behavioral abilities.

A more direct approach to dynamic aspects of behavior, however, is through the analysis of factors of motivation and emotion. Very much has already been done in this direction and Cattell (1957), in particular, has generated considerable theory from factorial information concerning needs, interests, and attitudes. The integration of such theory with that derived from intellectual factors is something still to be done.

SUMMARY

In explaining how factor analysis and its findings can contribute to general psychological theory, the writer described three models, derivable from factor theory or from the results of factor analysis, that apply to personality. Personality is defined as a matter of individual differences with respect to traits. One model is a space of n dimensions, each dimension representing a unique trait that is discoverable by factor analysis. Another model that emphasizes the structure of a typical personality represents traits at several distinct levels of generality, one of the levels being devoted to primary traits. A third model arises from the logical classification of known primary traits in columns and rows to form two- and three-dimensional matrices.

It is also necessary to discuss the adequacy of factor analysis as a method adaptable to the discovery of psychological concepts having theoretical significance. Used in an appropriate type

of experimental design and applied to the appropriate kinds of experimental variables, factor analysis can be a powerful tool for the meaningful discrimination among traits. Hypothesis testing of a kind is possible, in spite of inadequate statistical tests of significance.

Although factor analysis starts with data regarding individual differences in behavior and comes out with primary traits that also refer to modes of individual differences, we find it possible to deduce corresponding ways of functioning in behaving organisms. In this manner we see a bridge between factors and functions. The classes of factors, as derived from the third kind of model, suggest general concepts that seem to have theoretical significance for psychology.

From the organization of intellectual factors we derive the impression that the organism is an agent that acquires information, retains it, uses it in generating new information, and evaluates information in connection with any of these steps. Learning is the achieving of information, and the various intellectual abilities are abilities to learn. Similarly, problem solving rests upon certain combinations of primary abilities, depending upon the circumstances. Reinforcement or effect in learning is accounted for by the operations of evaluation. Drive reduction and effect supply information as bases for evaluation, along with other feedback information. Motives and emotions, too, may be brought under the heading of information that concerns the organism's internal states.

The concept of association is very inadequate in accounting for the kinds of products with which the intellectual abilities are concerned. The products, such as units, systems, transformations, and implications, have about them a totality aspect that calls for something like the concept of Gestalt. The factorial approach offers a form of analysis that should be useful to the Gestalt psychologist as well as to the functional psychologist.

Like behaviorism, the proposed factor psychology applies a completely objective approach and yields objective descriptions. Unlike behaviorism, it aims at the construction of a model of what goes on between stimulus and response rather than a model emphasizing stimulus-response correlations. Also unlike behaviorism it finds no use for the concept of association and would substitute the concept of cognition-response sequences for that of stimulus-response associations.

From a systematic standpoint, factor psychology is best described as a type of functional psychology, without any close affiliation with the traditional functional school. With further extensions of factor-analytical efforts in the area of dynamics of behavior, we should be able to develop a more complete system than it is possible to present today.

REFERENCES

BURT, C. The structure of mind: A review of the results of factor analysis. *Brit. J. Psychol.*, 1949, **19**, 100–111, 176–199.

CATTELL, R. B. *Personality and motivation structure and measurement.* New York: World Book, 1957.

EYSENCK, H. J. The organization of personality. *J. Pers.*, 1951, **20**, 101–117.

FERGUSON, G. A. On learning and human ability. *Canad. J. Psychol.*, 1954, **8**, 95–112.

FLEISHMAN, E. A., & HEMPEL, W. E., JR. Changes in factor structure of a complex psychomotor test as a function of practice. *Psychometrika*, 1954, **19**, 239–252.

FLEISHMAN, E. A., & HEMPEL, W. E., JR. The relation between abilities and improvement with practice in a visual discrimination reaction time task. *J. exp. Psychol.*, 1955, **49**, 301–312.

GUILFORD, J. P. Factors in problem solving. *ARTC Instructors J.*, 1954, **4**, 197–204.

GUILFORD, J. P. *Personality.* New York: McGraw-Hill, 1959. (a)

GUILFORD, J. P. Three faces of intellect. *Amer. Psychologist,* 1959, **14**, 469–479. (b)

GUTTMAN, L. A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences.* Glencoe: Free Press, 1954.

LAZARSFELD, P. F. A conceptual introduction to latent structure analysis. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences.* Glencoe: Free Press, 1954.

MAXWELL, A. E. Statistical methods in factor analysis. *Psychol. Bull.,* 1959, **56**, 228–235.

MERRIFIELD, P. R., GUILFORD, J. P., CHRISTENSEN, P. R., & FRICK, J. W. A factor-analytic study of problem solving. *U. Southern Calif. Rep. Psychol. Lab.,* 1960, No. 22.

MILLER, N. E. Shortcomings of food consumption as a measure of hunger: Results from other behavioral techniques. *Ann. NY Acad. Sci.* 1955, **63**, 141–143.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

SEWARD, J. P. Basic issues in learning theory. In G. S. Seward & J. P. Seward (Eds.), *Current psychological issues.* New York: Holt, 1958.

SPEARMAN, C. *The nature of 'intelligence' and principles of cognition.* New York: Macmillan, 1923.

SPEARMAN, C. *Abilities of man.* New York: Macmillan, 1927.

VERNON, P. E. *The structure of abilities.* New York: Wiley, 1950.

(Received December 2, 1959)

Psychological Review
1961, Vol. 68, No. 1, 21–32

PERSONALITY THEORY AND THE PROBLEM OF STABILITY AND CHANGE IN INDIVIDUAL BEHAVIOR:

AN INTERPERSONAL APPROACH [1]

PAUL F. SECORD AND CARL W. BACKMAN
University of Nevada

Clinical psychologists, personality theorists, and most other students of individual behavior commonly assume that a person gradually forms characteristic behavior patterns which become more and more resistant to change with the passage of time. These patterns are usually thought of as reflections of intra-individual structures or mechanisms: e.g., habits, needs, cognitive structures, or traits. Because these intrapersonal determinants of behavior are, in effect, accepted as givens, observable consistency over a period of time tends to remain unexplained—it is simply a manifestation of these structures. From this point of view, a person behaves as he does because of what he is.

Thus, Allport (1937), Cattell (1950), and Eysenck (1953) attribute enduring structure to personality, a structure in which the concept of trait is central. Murray (1938) conceives of personality as an organizing and integrating force having a neural locus in the organism. In spite of his field emphasis, Lewin (1935) conceives of personality structure as characterized by certain differentiations and articulations of regions, which correspond to aspects of personal character. Murphy (1947) includes physiological dispositions, canalizations, conditioned responses, and cognitive and perceptual habits as components of personality. Two cognitive orientations are those of Rogers (1951), and Combs and Snygg (1959), who conceive of behavior as organized around and guided by a relatively enduring self-concept. Finally, almost all forms of personality assessment are based upon the assumption that the individual undergoing appraisal has a stable personality structure which the investigator is attempting to describe, in part or in whole. Murphy (1947) has summed up this attitude succinctly. He states:

> The laboratory psychologist and the clinician have both conceived the individual as a system of events and tendencies carried around within the skin of the individual subject. They have both assumed that his delinquencies or his triumphs, his interests or nervous maladjustments, result fully and simply from structured dispositions within him (p. 877).

For all of these theorists, then, personality has a certain stability of structure which in turn maintains continuity in behavior over time. This is not to assert that behavior is viewed by these theorists as unchanging or unresponsive to varying stimulus situations. On the contrary, many of these personologists often give careful attention to stimulus situations as modifiers of behavior. But, sometimes explicitly and sometimes implicitly, they do appear to make the assumption that when behavioral stability occurs, it is a function of stability in personality structure. A closely related assumption is that personality structure has a strong resist-

[1] This investigation was supported in part by a grant (M-1892) from the National Institute of Mental Health, United States Public Health Service.

ance to change; in the absence of special change-inducing forces or conditions, it remains constant.

These assumptions have two consequences. The first is that continuity in individual behavior is not a problem to be solved; it is simply a natural outcome of the formation of stable structure. The second is that either behavioral change is not given systematic attention, or change is explained independently of stability. Whereas behavioral stability is explained by constancy of structure, change tends to be explained by environmental forces and fortuitous circumstances. A more parsimonious approach would be to account for both stability and change by means of a single set of explanatory principles. The following theoretical statement abandons the assumption that internal stabilizing mechanisms are inherently resistant to change and are the sole source of stability in individual behavior.

The present paper places the locus of stability and change in the interaction process rather than in intrapersonal structures. At the same time, the importance of habit, cognitive structure, and similar concepts is not denied, but these mechanisms do not by themselves establish necessary conditions for behavioral stability. Stable patterns in the interaction process, which constitute the essential condition for consistency in individual behavior, may be thought of as deriving from two major sources. One source is the cultural, normative, and institutional forces which stabilize the behavior of persons with whom a particular individual interacts, as well as his behavior toward them. With this important source the present paper is not concerned; social psychologists, sociologists, anthropologists, and others have analyzed these forces in extensive detail.

The other source of stability, although it has been given some attention by social psychologists, has been relatively neglected by students of individual behavior. It has been customary to think of two classes of behavioral determinants: cultural and normative forces, and those forces stemming from the individual. The present paper attempts to identify a third class of determinants, which have their locus neither in the individual nor the culture, but in the interaction process itself. In a general sense this third class may be characterized as the tendencies of the individual and the persons with whom he interacts to shape the interaction process according to certain requirements, i.e., they strive to produce certain patterned relations. As will be seen, the principles governing this activity are truly interpersonal; they require as much attention to the behavior of the other as they do to the behavior of the individual, and it cannot be said that one or the other is the sole locus of cause.

Because of limitations of space and in the interests of clarity, the theory will be presented without attempting to specifically examine the empirical evidence, but those familiar with the literature will recognize that much supporting evidence exists. For definitive support, however, an extensive research program guided by the theory needs to be carried out.

Sources of Stability in Individual Behavior

Definition of Interpersonal Matrix and Matrix Congruency

The locus of behavioral stability and change lies in the interpersonal matrix, which has three components: an aspect of the self-concept of the subject (S), S's interpretation of those elements of his behavior related to that aspect, and S's perception of related aspects of the

other person (O) with whom he is interacting. An interpersonal matrix is a recurring functional relation between these three components. The behavior of a particular person is episodic in character, in that he shifts from one matrix to another over varying periods of time. A matrix is referred to as the same if, on two or more occasions, the same self-component, S-behavior, and O-component are present in the same functional relation. The present theoretical treatment focuses mainly on the discussion of single matrices. Also, only aspects of self and behavior which are valued by *S* are dealt with, and these are assumed to maintain relatively constant values.

S strives to achieve congruency[2] among the components of the matrix. Congruency is a cognitive phenomenon: i.e., each component enters into a state of congruency only as a perceptual cognitive experience on the part of *S*. All three components of the matrix are in a state of congruency when the behaviors of *S* and O imply definitions of self congruent with relevant aspects of the self-concept. In a nurturant-dependent relation, for example, a dependent *S* has conceptions of self containing elements of inadequacy and need for support. His behavior is consistent with these aspects of cognitive structure, for it is hesitant and compliant, reaffirming his self-definition. Since the behavior of O is perceived as providing assistance and as reassuring and decisive, it implies a matching definition of *S*'s self-concept and behavior. In the above example, *S*'s interpretation of his behavior is congruent with a relevant aspect of the self, *S*'s perception of O is congruent with the self-aspect, and *S*'s interpretation of his behavior is congruent with his perception of O. Thus, any two components of a matrix may be congruent with each other, with the third either congruent with both of them, or not congruent with either.

Implications for self-definitions may take three forms: *S* may perceive O's behavior as directly confirming a component of self, O's behavior may enable *S* to behave in ways that would confirm a component of self, O's behavior may (by comparison) lead other Os to confirm a component of *S*'s self-concept. Examples of each form are:

An *S* who regards himself as mature and responsible perceives that Os respect him for these characteristics.

An *S* who regards himself as nurturant encounters an O in need of help; this allows him to behave toward O in a manner which supports his nurturant aspect of self.

A girl who regards herself as popular and well-liked keeps company with an unpopular girl; Os are viewed by her as judging her favorably by contrast.

Reality-oriented vs. cognitively distorted matrices. A matrix is reality-oriented if the S-behavior component and O-component are correctly interpreted or perceived. On the other hand, a matrix which involves misinterpretation of *S*'s behavior or misperception of O is not reality-oriented; it involves some form of cognitive distortion. It is obvious that congruency would sometimes be achieved by means of cognitive distortion. Presumably a congruency achieved by this means would be less stable than a reality-oriented congruency. At least this would be true if there were forces tending to produce accurate interpretations and perceptions. On the other hand, *in*congruencies which might be made congruent simply by correcting cognitive distortions to fit reality are not

[2] Congruency as used here has some similarities to the following: Abelson's (1958) resolution of belief dilemmas, Cartwright and Harary's (1956) concept of structure balance, Festinger's (1957) theory of cognitive dissonance, Heider's (1958) theory of balance, and Osgood and Tannenbaum's (1955) concept of congruence.

regarded as particularly stable or significant, and thus are not further discussed.

The tendency to achieve and maintain congruent states. A number of general principles may be presented concerning the maintenance of congruent states. In many instances these will be elaborated later.

1. *S* tends to repeat and perpetuate those interpersonal relations which were previously characterized by congruency.

2. An *S* involved in a matrix which is not in a state of congruency will tend to modify the matrix in the direction of greater congruency.

3. The engagement of *S* and O in congruent interaction develops mutual affect toward each other, which tends to perpetuate the interaction.

4. Because of the tendency to establish congruent matrices, *S* gradually builds up an increasingly greater repertory of such matrices.

5. The more the O-component of a congruent matrix is valued, the greater the tendency of the matrix to be perpetuated.

Relations among matrices and the perpetuation of congruent states. A given matrix may be considered *relevant* to those matrices which contain one or more of the same or similar components as the given matrix. For example, for an *S* who considers himself to be a kind person, there will be many S-behaviors and O-behaviors which are perceived as forming congruencies and inconguencies with this aspect of self. The various matrices formed by distinctive S-behaviors and O-behaviors in connection with the aspect of self, kindness, are all considered relevant to each other.

A relevant matrix may be either *supportive* or *nonsupportive* with respect to a given matrix. Supportiveness between two matrices is determined by whether or not the componont(s) common to them has the same relation (one of congruency or one of incongruency) to the other unique components in each matrix. If the common component has the same relation in both matrices, they are said to be supportive.

For example, suppose a woman regards herself as unattractive to men. Her mother continually calls her attention to the fact that women in her family were seldom noted for their attractiveness. Our *S* also finds that a man she likes at work avoids her as much as possible. In the one matrix, unattractiveness is congruent with her mother's direct definition of her as unattractive and, in the other, the avoidance behavior of the man leads her to perceive his behavior as defining her as unattractive. Thus, one matrix supports the other. On the other hand, if her mother perceived many good features in her appearance and stressed these, and the man professed admiration for her, both matrices would be incongruent with her notion of self as unattractive. These two matrices would also support each other—they agree in denying an aspect of self. These might be termed incongruent supportive matrices, and the former type, congruent supportive matrices.

A statement may now be made with respect to the role of a complex of matrices in perpetuating a given matrix. Matrices vary with respect to *centricality*. The centricality of a matrix is a function of the number of other matrices which stand in a supportive relation to it, and the value of the O-components in these matrices. The greater the centricality of a matrix, the more resistant it is to change, and, should it change, the greater the resultant shifts in other matrices.

Interpersonal Processes Contributing to Congruent States

Congruency is continually threatened by the changing nature of interpersonal relations due to: (*a*) normative patterns of change, which result in changes in the behavior of others towards *S*,

e.g., changes in the behavior of others as the child grows older, or as a person changes occupations, gets married; (*b*) fortuitous changes, e.g., death of a parent, loss of a friend, induction into the military; and (*c*) the fact that O's means of establishing congruency for himself often create incongruencies for *S*. A number of interpersonal processes may be suggested which counter these pressures for change and which perpetuate previous matrices.

The first five interpersonal processes restore congruency through some transformation of the O-component of the matrix:

1. *Selective interaction with Os.* *S* tends to maximize engagement in congruent patterns of interpersonal behavior by selecting and interacting with those Os whose behavior requires a minimum change from previously congruent interpersonal situations in which *S* has engaged. For example, if *S* regards himself as especially intelligent, he tends to interact frequently with Os who respect his intelligence or who allow him to exercise it, or he interacts in such a way as to make himself appear intelligent by comparison.

2. *Selective evaluation of Os.* *S* tends to maximize congruency by altering the evaluation of selected Os in a positive or negative direction, depending upon whether they are behaving congruently or incongruently with certain aspects of self. The more an O is valued by *S*, either positively or negatively, the more important his role is in maintaining congruency. This provides a means of increasing the effects of congruency or reducing the effects of incongruency. Thus, *S* tends to increase his liking for Os who behave toward him in a congruent fashion, and to decrease his liking for those who behave in an incongruent manner.

3. *Selective comparison with aspects of O.* *S* tends to maximize congruency by selectively attending to behaviors of O which are congruent, and selectively ignoring those which are incongruent. For example, a patient in therapy particularly notes those behaviors on the part of the therapist which confirm aspects of his self-concept and behavior, and tends to ignore those which are incongruent.

4. *Evocation of congruent responses from O.* *S* develops techniques for eliciting from O behavior which will be congruent with components of his self-concept and behavior. For example, a feminine woman plays a helpless role in order to elicit dominance from men.

5. *Misperception of O.* *S* may misperceive O's behavior so as to achieve congruency with aspects of his behavior and self-concept. Thus, an *S* has a high regard for himself as a lover and perceives other women as more interested in him than they really are.

The sixth and seventh interpersonal processes restore congruency by a transformation of the S-behavior component of the matrix:

6. *Selective behavior-matching.* In interacting with a particular O, *S* tends to maximize congruency by selecting from his total behavioral repertory those behaviors which are most congruent with his perception of O. This principle reflects William James' (1890, Vol. 1, p. 294) notion that each person has as many "social selves" as there are classes of persons whose opinion he cares about.

7. *Misinterpretation of own behavior.* *S* may misinterpret his behavior so as to achieve maximum congruency with an aspect of his self-concept and his perception of O. For example, a dutiful mother spanks her child because of anger and frustration, but convinces herself that she is doing it for the child's own good.

Certain combinations of these principles may sometimes operate under a single rubric. For example, an individual may select a *social role* which enables him to achieve maximum congruency among the three components. This involves interaction with selected Os who will engage in certain desired reciprocal behavior, and also permits behaviors which validate the self. Thus, a woman who regards herself as intelligent, independent, and ambitious may reject the marital role in favor of a career role. A second example is found in *circular interaction systems*. In these, response evocation tends to become an activity mutual to two persons: the actions of *S* give rise to behavior on the part of O which in turn evoke further actions of the type initially displayed by the individual. For example, aggressive actions on the part of the mother, initiated in the attempt to control or eliminate undesirable aggressive behavior in her child, raise his frustration level, which in turn gives further impetus to his tendency to respond in aggressive ways.

FORCES TOWARD CHANGE AND THE RESOLUTION OF THESE FORCES

States of Incongruency and Their Consequences

Any theory of stability in self and behavior must also be able to explain change. There are three steps leading to change: the creation of an incongruency, the formation of a new congruent matrix which involves a different component of self or behavior from that existing prior to the change, and the adjustment of relevant matrices which have been affected by the changes made in resolving the incongruent matrix. As mentioned earlier, incongruencies arise from several sources: from cultural, normative, and institutional forces, from fortuitous factors, and from O's attempts to establish congruency for himself. Incongruencies occur frequently, but are often followed by restoration of a congruent matrix which does not involve a change in self or behavior. Whether or not change occurs, resolution of the newly formed incongruency occurs through utilization of one or more of the interpersonal processes previously discussed. Thus, a single set of interpersonal processes are presumed to underlie both behavioral stability and the direction of behavioral change.

Four types of incongruency are identified in the following discussion. Of considerable interest and importance is the manner in which the various types of incongruency may be resolved. As already noted, there are two general classes of resolution. One of these results in restoration of the original matrix, leaving self and behavior unchanged (although cognitive distortions may occur), and the other leads to a new matrix in which self or behavior are changed. In order to avoid being tedious, only a brief, informal discussion of the forms of resolution will be presented. Table 1, however, presents in symbolic form a complete set of resolutions for the first three types of incongruency. The fourth is omitted since it is mainly of academic interest and is probably quite rare.

Type I. The behavior of O is perceived as incongruent with a component of self and with *S*'s behavior, the latter two elements being congruent with each other.

For example, *S*, an alcoholic, who regards himself as a "weak," inadequate person, neglects his family by going on extensive binges, implying a definition of self as inadequate. O, his wife, behaves toward *S* as if he were capable of controlling his drinking.

In this type of incongruency, *S* may employ any one of the first five interpersonal processes to transform the incongruent O-

TABLE 1

RESOLUTION OF BASIC TYPES OF INCONGRUENCY

Let E represent the self-component of matrix
B represent S-behavior as a reality-oriented matrix component
O represent O-behavior as a reality-oriented matrix component
B′ represent a matrix component based on misinterpretation of S-behavior
O′ represent a matrix component based on misperception of O-behavior
$\bar{X}$ represent an incongruent form of a component X
X_c represent a change in a component X

Type of Incongruency	Type of Resolution: Maintenance of Self and Behavior	Change in Self-Concept	Change in Behavior	Change in Self-Concept and Behavior
I. $SB\bar{O}$	SB SB SBO_e SBO_c SBO'	$S_cB'O$	Not applicable	S_cB_eO
II. $\bar{S}BO$	SB′ SB′ $SB'O_e$ $SB'O_c$ $SB'O'$	S_cBO	SB_e SB_e SB_eO_e SB_eO_c SB_eO'	Not applicable
III. $S\bar{B}O$	SB′O	S_cB S_cB S_cBO_e S_cBO_c S_cBO'	SB_eO	Not applicable

Note.—Resolutions involving O are listed in order of the five interpersonal processes which apply to O: (*a*) reduced interaction—SB, (*b*) devaluation—SB, (*c*) O-behavior selection—SBO_e, (*d*) response evocation—SBO_e, and (*e*) misperception—SBO′.

component, and thereby maintain self and behavior unchanged. Thus, he may reduce interaction with his wife, may devaluate her, may avoid selective comparison by declaring her opinion irrelevant, may evoke new responses from her, or may falsely perceive her as trying to give him a pep talk.

Other forms of achieving congruency require a change in self or behavior. *S* may change both components, or he may change self but not behavior by misinterpreting his behavior.

Type II. An aspect of self is incongruent with *S*'s behavior and with the perceived behavior of O, the latter two elements being congruent with each other.

For example, a woman behaving seductively gets matching responses from O, but regards herself as morally conservative.

One form of resolution involves maintenance of self and behavior. Since behavior is incongruent, it would have to be misinterpreted if self is to be maintained, and in addition, the five interpersonal processes which transform the incongruent O component would be employed here. For example, behavior is misinterpreted if *S* regards her behavior as casually flirtatious, or as "not me" because she had been drinking, and, with respect to O, *S* might reduce interaction with him, regard his behavior as harmless flirtation, or she might make her own behavior irrelevant by defining him as a man who would try to seduce any woman, etc.

A simple change in self would result in a congruent matrix. Or, finally, *S* might change her behavior and transform the O-component.

Type III. *S*'s behavior is incongruent with a component of self and with the perceived behavior of O, the latter two elements being congruent with each other.

For example, a wife regards herself as loving, but realizes that she is often hostile toward her husband. Her husband fails to recognize much of the hostility and behaves toward her as if he perceives her as loving.

If, in this instance, *S* changes her self-concept to match her unchanged behavior, she may employ the five interpersonal processes to transform the O-component. Thus, she may leave her husband, devaluate him, emphasize occasions on which he has been angry with her, evoke hostility from him, or think that he secretly does recognize her hostility. If she maintains self-concept and behavior, she must necessarily misinterpret her behavior in order to achieve congruency.

Type IV. All three matrix components are incongruent with each other. Two general forms of resolution are: (*a*) any two of the components may be changed to match the third; (*b*) all three may be modified to achieve a new matrix congruency. Type IV incongruency is thought to be rare and is not discussed further.

Factors Governing the Probable Direction of Resolution

The various forms of resolution of incongruent matrices are not equally probable. Which ones are more likely to occur is in part a matter for empirical study. In the meantime, however, a number of factors relating to the probability of change in the various components may be identified. In the first place, O's behavior toward *S* may be systematic and consistent because it stems from certain factors external to the dyad, such as role expectations—e.g., O may be *S*'s therapist, or he may be a parent socializing his child. An obvious second consideration is the ease with which *S* may leave the dyad. A patient may readily leave therapy, but a child cannot usually leave his family. A third condition pertains to the extent to which an individual's congruent relations are reality-oriented, i.e., to what degree misinterpretation and misperception are required for congruency. Matrix components which are not reality-oriented tend to be less stable. A fourth point is that the greater the number and value of supportive congruent matrices which are relevant to a component, the more resistant is that component to change. Finally, it may be postulated that the individual learns to utilize selectively the various modes of resolution. For a particular individual, certain types of component-transformation are more practiced and are preferred to others.

DISCUSSION

An important difference between the present approach and earlier ones should be re-emphasized. Such traditional concepts as cognitive structure, self-concept, trait, and habit are products of earlier social interaction. But the assumption is *not* made that persons strive to maintain either their self-concept or various habitual behaviors by virtue of some inherent "gyroscopic" force residing in these dispositions. Maintenance of intrapersonal structure occurs only when such maintenance is consistent with an ongoing interaction process which is in a state of congruency. That most individuals do maintain intrapersonal structure is a function of the fact that the behavior of others toward the individuals in question is normally overwhelmingly consistent with such maintenance. This fact has often been overlooked. It should be made clear that the concept of structure continues to have a place in the present theory. Matrices are structures; they are aspects of perceptual-cognitive experience. Unlike habit and trait, however, they tend to be more integrated with certain behaviors of O. Even the self-concept, employed in the present theory as a matrix component, is always tied to certain behaviors of *S* and O.

The interpersonal environment is not always stable and familiar. The present approach predicts that were the interpersonal environment to suddenly undergo drastic change, with others uniformly behaving toward the individual in markedly new and strange ways, an individual held in such an environment would rapidly modify his own behavior and internal structure to produce a new set of congruent matrices. As a result, he would be a radically changed person. This prediction is consistent with the experiences of some soldiers imprisoned and "brainwashed" by Chinese communists. Some investigators have emphasized as a key factor in successful brainwashing the isolation of the individual from his normal interpersonal relations. Marked changes in an individual may also occur as a result of shifts in a single matrix, if that matrix involves a highly valued O, e.g., a therapist.

As noted earlier, there are complexes consisting of matrices relevant to each other because they contain at least one common component, and a shift in one highly valued matrix may have ramifications throughout the structure of relevant matrices. In a word, the differences in emphasis between the present and previous approaches to behavioral stability is that the individual strives to maintain interpersonal relations characterized by congruent matrices, rather than to maintain a self, habits, or traits. Some of the advantages of this approach are the following:

1. The approach provides a place in theory for the variability of individual behavior in different interpersonal situations. For example, it does not make the questionable assumption that an individual belongs at fixed points with respect to each of a variety of personality traits in all interpersonal situations. A given individual is unlikely to be equally aggressive in interaction with every other person or in every situation involving the same person. Yet, a typical personality appraisal instrument yields a single score on "aggressive" for an individual. In the present view, trait behaviors such as aggressiveness tend to appear only when they are congruent with certain O-behaviors and certain components of self. This idea calls for an overhaul of our traditional methods of personality assessment.

2. While self-theory emphasizes that self emerges out of social interaction, it is rather vague on details. In fact it implies that the only role of the behavior of others in supporting the self is that expressed by the first type of congruent O-behavior, namely, behavior which directly defines an *S*-component of *S* in the same way as *S* does. Moreover, the organism is often assumed to be passive in this process. For example, the fact that *S* may actively seek out those who support components of self, may evoke certain responses on the part of the other, etc. is often ignored.

3. The interpersonal framework makes the process of individual change less mysterious and less difficult to explain. Individual behavior is always linked to the behavior of others, and the key to deliberate change lies in creating shifts in the perception of the behavior of Os toward *S*. This is accomplished either through bringing about actual changes in the behavior of Os toward *S*, or by blocking various techniques that *S* uses to distort or to evoke certain O-behaviors which tend to support his current perceptions. Socialization and psychotherapy will shortly be discussed in order to illustrate the application of the theory to problems of change.

4. It is believed that variables on the perceptual-cognitive level are more directly assessable in view of the ability of *S*s to directly describe their self-concept, their behavior, and the behavior of Os. In addition, the question of cognitive distortion can be checked by obtaining information by direct observation of the interaction, or by obtaining descriptive information from O. In contrast, such concepts as need, habit, and trait must be inferred, and there has been considerable controversy over the validity of various procedures of inference in regard to these variables. Another advantage stems from the fact that, in the experience of the present investigators, matrix variables are readily manipulable when using conventional experimental procedures in social psychology. For example, an incongruency can be readily created in the laboratory by false test protocol reports to *S*s concerning aspects of self and behavior, or by placing *S* in the laboratory with an O who has been

instructed by the experimenter to behave in an incongruent fashion.

5. The present point of view should be helpful in integrating concepts of personality with those of social psychology. At present there is a tendency to conceptualize each of these disciplines in such a way as to make difficult, if not impossible, the simultaneous application of concepts from both areas to consideration of any particular behavioral phenomenon. Dyadic relations will be briefly touched upon to indicate how a rapprochement between these areas might be initiated. Also, by examining the role of culture in establishing congruencies and incongruencies, light may be thrown on variations in individual strain from culture to culture, and on a source of cultural change, namely, the attempt of individuals to resolve such strains.

Socialization and psychotherapy. The process of socialization involves those changes occurring as a result of movement through the social structure. Psychotherapy has a less systematic relation to social structure, and may be thought of as a process of change resulting from highly significant interaction between patient and therapist. Both socialization and psychotherapy may be examined briefly to determine whether or not the process of change can be conceptualized in terms of the present interpersonal theory.

Socialization may be viewed as a process in which the person moves through a succession of changing patterns of O-behaviors over a long period of time. These changing patterns are culturally induced, and create incongruencies for *S*. In this process there is a certain amount of resistance to change on the part of *S*, arising out of the fact that typically the changing behavior of Os toward him requires continual reorganization of previously established matrices. Behaviors which previously led readily to congruent matrices now bring forth incongruent reactions from O. Take, for example, the small child who has an overprotective mother. Here self-cognitions are rudimentary and limited, but if existing components were phrased in adult language, they might include the following: "I am the baby of the family." "Daily family life revolves around me." "I can get most anything I want." When this child moves out of his mother's sphere of control, however, and plays with other children, he meets many responses incongruent with these components of self. Other children are likely to thwart his needs rather than meet them; he is unlikely to be the center of play activities. These incongruencies force change in self-components and behavior. To the extent that an overprotective mother limits a child's contacts with other children, and reduces the extent to which other children are valued, components of self inappropriate to interaction outside the family develop in the child.

Normal development in the socialization process involves removing incongruency by changing components of self and behavior so that they correspond with the definitions made by significant Os. Because of the difficulty of changing S-components which are relevant to an increasingly larger number of congruent matrices, however, the individual also learns to use other ways of handling incongruency: e.g., misperception, misinterpretation, response evocation, changing valuations of others, etc. These latter actions delay or prevent changes in self or behavior. It is probable that early adjustment to incongruency involves changing the self-concept and behavior, but as the individual develops these other techniques, he frequently handles incongruency by means of them. Disturbances of the developmental se-

quence may frequently occur as a result of overuse of modes of reducing incongruency which involve distortion of reality instead of change in self or behavior. This may be expected to result in the establishment of a somewhat precarious stability of self at a low level of maturity. This is a social psychological interpretation of the Freudian phenomenon of fixation.

The interpersonal theory as outlined also seems to point to a crucial process in psychotherapy. The modes of reducing incongruency spelled out in an earlier section suggest that the primary problem facing a psychotherapist is the necessity for blocking all modes which permit the maintenance of the undesirable aspects of self and behavior, and facilitating those which involve change in these components. The advantage of a systematic theory providing an exhaustive classification of the various forms of reducing incongruency is that it allows the therapist to anticipate various ways in which the individual may fail to change. There are two directions, however, in which the theory must be further elaborated. One involves spelling out the behaviors required of the change agent for successful facilitation of modes of resolution leading to desired changes in the self-concept and behavior of the *S*. The other requires the elaboration of techniques for blocking those modes leading to maintenance of undesirable aspects of self and behavior.

Dyadic relations. So far the focus has been on stability and change in *individual* behavior. Emphasis has been placed on the point that stability and change in individual behavior is in part a product of a reciprocal process having its locus in the interactions between the individual and those around him. This reciprocal process may also be applied to an understanding of both members of the dyad. The instability of adolescent love relations would appear to be initiated by the confronting of the adolescent with marked and frequent changes in the behavior of peers and adults which are incongruent with components of self and his behavior. These changes lead to shifts in self-concept and behavior, and in turn new relations are sought to establish congruency with these modified aspects of self and behavior. The disruptive effect of separation on a friendship or a marriage is another case in point. The new interpersonal environment contains Os who define the individual's self and behavior differently, and he gradually changes so as to establish congruency in this new interpersonal situation. On the other hand, old relations become less congruent, and consequently progressively weaker in affect. These changes are dramatically illustrated by the disappointment so often experienced at the reunion after long separation of boyhood friends or ex-army buddies. In a similar fashion, induced changes in personality, such as those which occur in psychotherapy, and fortuitous ones which arise as a result of entering a new job, obtaining more formal education, etc. may well result in drastic shifts in dyadic relations.

SUMMARY AND CONCLUSIONS

Most students of individual behavior commonly assume that when behavioral stability occurs, it is a function of stability in personality structure, and that in the absence of special change-inducing forces, structure remains constant. Thus continuity in individual behavior is not a problem to be solved; it is simply a natural outcome of the formation of stable structure. Behavioral change, moreover, either is not given systematic attention, or is explained independently of stability, in terms of environmental forces and fortuitous

circumstances. The present interpersonal theory attempts to account for both stability and change in terms of a single set of explanatory principles.

In the present view, the locus of behavioral stability and change lies in the interpersonal matrix, which has three components: an aspect of the self-concept of *S*, *S*'s interpretation of those elements of his behavior related to that aspect, and *S*'s perception of related aspects of the person with whom he is interacting (O). *S* strives to achieve congruency among the components of the matrix, a state which is achieved when the behaviors of *S* and the other imply definitions of self which are consistent with relevant aspects of the self-concept. Congruency is continually threatened by the changing nature of interpersonal relations due to: (*a*) normative patterns of change, which result in changes in the behavior of Os toward *S*, e.g., changes in the behavior of Os as the child grows older, or as a person changes occupations or gets married; (*b*) fortuitous changes, e.g., death of a parent, loss of a friend, induction into the military; and (*c*) the fact that O's means of establishing congruency for himself often create incongruencies for *S*.

When incongruencies arise, a variety of interpersonal processes which have been carefully described here often act to restore congruency without change in self or behavior. Sometimes, however, these same interpersonal processes operate to form a new congruent matrix which involves a change in self or behavior. The circumstances under which either of these effects take place are a function of the type of incongruency which occurred and of a variety of other conditions which have been discussed in some detail. Thus, a single set of principles are utilized to offer a systematic explanation of the reasons for stability or change in individual behavior.

The advantages of adopting such an interpersonal theory are spelled out, and a brief discussion is included to illustrate how it might be applied to such processes as socialization, psychotherapy, and dyadic relations.

REFERENCES

ABELSON, R. P. Modes of resolution of belief dilemmas. Paper read at Western Psychological Association, Monterey, California, April 1958.

ALLPORT, G. W. *Personality: A psychological interpretation.* New York: Holt, 1937.

CARTWRIGHT, D., & HARARY, F. Structural balance: A generalization of Heider's theory. *Psychol. Rev.*, 1956, **63**, 277–293.

CATTELL, R. B. *Personality: A systematic, theoretical, and factual study.* New York: McGraw-Hill, 1950.

COMBS, A. W., & SNYGG, D. *Individual behavior: A perceptual approach to behavior.* New York: Harper, 1959.

EYSENCK, H. J. *The structure of human personality.* New York: Wiley, 1953.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson, 1957.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

JAMES, W. *Principles of psychology.* New York: Holt, 1890. 2 vols.

LEWIN, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.

MURPHY, G. *Personality: A biosocial approach to origins and structure.* New York: Harper, 1947.

MURRAY, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.

OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychol. Rev.*, 1955, **62**, 42–55.

ROGERS, C. R. *Client-centered therapy: Its current practice, implications, and theory.* Boston: Houghton Mifflin, 1951.

(Received January 16, 1959)

Psychological Review
1961, Vol. 68, No. 1, 33–45

REALISM OF CONFIDENCE JUDGMENTS [1]

JOE K. ADAMS [2]
Palo Alto Medical Research Foundation

AND

PAULINE AUSTIN ADAMS [3]
Menlo Park, California

The problem of the validity of beliefs has occupied Western philosophers from the time of the early Greeks on down to the present day. The modern form of this problem is that of finding criteria according to which *degrees* of belief can be justifiably assigned to assertions (or their negations). Psychologists, economists, mathematicians, and others have also been concerned with this question, usually in a less general way, i.e., they have usually been concerned with beliefs of more or less specific kinds. There is very limited agreement at present either on the *normative* problem of finding adequate criteria for various degrees of belief or on the more *empirical* problem of measurement of degrees of belief. Within psychology the normative problem has been attacked primarily on the level of finding adequate criteria for holding some degree of belief in an assertion made professionally, i.e., on the level of methodology. The measurement problem has been attacked in several lines of research, two of which come under the headings of "confidence" and "subjective probability." Measurement of confidence has usually been of an admittedly ordinal type, obtained by direct judgments by the subject (S), whereas measurement of subjective probability has usually been of an assertedly interval- or ratio-type obtained by inference from betting decisions. Except for very restricted classes of events, adequate criteria for realism of confidence or subjective probability have been lacking. In what follows we shall expand these remarks and propose a method of measuring confidence (or subjective probability) which is both descriptive and, in a sense, normative, in that the method can be used to obtain a meaningful measure of realism in judgment, for any of a wide range of classes of events. We shall also describe several areas of application of this method.

Approaches to the Problems Involved in Confidence or Subjective Probability

Epistemological discussions have traditionally been concerned with the general question of whether *any* of our beliefs can be justified; [4] recently, however, a much greater emphasis has been put on the problem of the *degree* of justification that can be adduced for a given assertion. As with any philosophical problem, wording rapidly becomes critical, and even to state the

[1] This research has been supported by Public Health Service Research Grant M-2297 and by Public Health Service Fellowship MF-6977. Grateful acknowledgment is made to E. R. Hilgard for several very helpful comments on a first draft of this manuscript.

[2] Formerly Research Associate, Applied Mathematics and Statistics Laboratory, Stanford University.

[3] Formerly Research Associate, Department of Psychology, Stanford University, and Research Fellow, USPHS.

[4] Three classical examples of philosophical questioning are the discussions concerning belief in the existence of external objects, belief in the existence of another person's experience, and belief in the validity of one's own memory. As has been noted by many others, to express skepticism about these beliefs in any way other than intellectual discussion can easily result in behavior that is odd, to say the least.

problem is to invite both confusion and a challenge. At least from one point of view, however, the ill-defined field of epistemology can be said to be concerned with finding some set of principles which enables us to assign, in some justifiable way, a degree of confidence to any assertion contained within some class of assertions with which we are concerned.

There have been noteworthy attempts to develop principles of great generality, which hold over a large range of assertions, or even over all. As in many analogous situations, however, there is an inverse relationship between generality and the extent to which the principles can actually be used. Carnap's system (1950), for example, while of extraordinary theoretical generality, cannot be used within the foreseeable future except with the relatively very limited classes of assertions covered by much less general systems. According to Georgescu-Roegen (1958), Carnap's work has made clear the impressive difficulties facing any system of comparable generality.

With more limited classes of assertions, various specialists have attempted to develop methods of evaluation which enable one to attach a justifiable degree of confidence, at least approximately, to each assertion within their particular field. In psychology, for example, attempts have been made to develop methods for evaluating assertions such as:

> For a single response, the greater the delay in reinforcement, the weaker the reaction potential, according to a gradient that falls off rapidly at first and then more slowly, approaching the asymptote at about 5 seconds (Hull, 1943, p. 243). Repression demands a constant expenditure of energy, and if this were discontinued the success of the repression would be jeopardized, so that a fresh act of repression would be necessary (Freud, 1925, p. 89).

These examples are chosen to point up the incomplete systematization and the disagreement among psychologists as to these methods—without any intended implication that this state of affairs is necessarily detrimental to the development of psychology. Nevertheless, we do have some systematization and agreement, which sometimes become apparent when we talk with specialists from other fields, who of course have their own more-or-less-developed-and-agreed-upon systems.

There exist at least two large classes of assertions for which there are well systematized and agreed upon methods for arriving at convictions. One of these is the class of mathematical propositions for which alleged proofs have been published—though even here both systematization and agreement are not entirely complete. For unproved mathematical propositions, however, the situation is quite different; for example, there is no known way of saying how confident one should be that Goldbach's conjecture (that for every even number greater than 2, two primes exist such that their sum is that number) is correct, despite the existence of some partial results and despite the fact that many mathematicians strongly believe this conjecture—perhaps in spite of themselves. There is a dichotomous aspect of justifiable belief in a mathematical proposition: either it has been acceptably proved and one can believe it with certainty, or one cannot assign any justifiable degree of belief—for example, there is no generally acceptable way of justifying or discrediting the statement that Goldbach's conjecture should neither be believed nor disbelieved, or that the number 0.5 should be assigned as the probability that it is correct.

The other large class of assertions for which there exist well systematized and agreed upon methods for arriving

at convictions is the class of propositions amenable to statistical methods. Unlike mathematical propositions, to these statistically amenable propositions we can attach *degrees* of conviction which are neither 0 nor 1. It is important to recognize the very limited range of these propositions, however. Such a proposition as that the difference in means between Condition A and Condition B, in all their concrete totality, did not "just happen" through random processes should not be confused with the proposition that a particular aspect (or combination of aspects) of Condition A (which is, after all, what we as psychologists are ordinarily interested in) results in a higher mean than a particular aspect of Condition B. For the identification of variables which are responsible for whatever we observe, we must rely on our much less systematized and agreed upon methods of observation and judgments based upon knowledge, experience, and theoretical convictions.

In our discussion we have long since passed the point of agreement among experts in these matters. There are almost as many views as to the relation of formal theory—as in logic, mathematics, and statistical methods—to belief or conviction as there are outstanding theorists. To relate formal proof to certainty, as in the foregoing, is unacceptable to some mathematicians and logicians; to identify the numbers (p values) which we obtain in statistical computations with a degree of conviction is even more controversial among those concerned with the foundations of probability theory and its applications. Among psychologists and other empirical scientists, however, there is, we believe, a strong relationship between p values and convictions that specific experimental results cannot be accounted for by random processes alone, granted that some regard the .05 level as quite satisfactory under conditions in which others demand that p be no greater than .01. For a discussion of the thorny problems of subjective interpretations of statistics the reader is referred to Savage (1954) and Georgescu-Roegen (1958); both authors give numerous references to works on the foundations of probability.

Quite aside from the normative problem of methods for assigning various degrees of conviction to assertions, even the measurement of the degree of conviction with which a given assertion is made, justifiably or not, by a given individual, is surrounded with difficulties, at least in the hands of some authors. We are referring especially to those assertions having to do with future events, such as the results of random processes (as in gambling) or events such as the increase or decrease in the market price of stocks or commodities, etc. Economists especially have been concerned with these events, and they recognized long ago that the behavior of an individual in a market situation is dependent, not upon objective probabilities (even assuming these exist, contrary to the strongly held convictions of some experts) and monetary values, but upon expectations (including subjective probabilities) and utilities. The economists took overt behavior in vivo as their prediction goal, and were apparently uninterested in the possible use of direct verbal estimates of probabilities and in the development of rating scales. The dictum that verbalizations of expectations usually have little to do with behavior seems to have been rather well accepted by economists and others working in the field of utility theory and subjective probability. Betting behavior, on the other hand, was accepted as indicating something about subjective probabilities, though it was generally agreed that, as subjective proba-

bility and utility work jointly, one could not be measured without assumptions about the other (Davidson & Suppes, 1957). von Neumann and Morgenstern (1947), for example, assumed that subjective probability is equal to objective probability; quite aside from the tenuous nature of this assumption it is not very helpful, because in most practical situations objective probabilities cannot be clearly defined, much less computed. Recently, however, Davidson and Suppes (1957) devised a method of measuring both utility and subjective probability, in a betting situation.

Psychologists have been more sanguine about the use of verbalizations of convictions, and have published many studies which include ratings indicating the degree of confidence with which various decisions have been made. That these ratings have some validity is shown by the frequent finding of a positive monotonic relationship between confidence (indicated on a verbal or numerical ordinal scale) and some measure of adequacy of performance. This relationship implies some validity in the confidence rating not only as a measure of conviction, but also as an index of the validity of the conviction. Possibly the earliest result of this kind is that reported by Peirce and Jastrow in 1884. The usual finding has been, however, that whereas *within* a given individual's judgments there is a high relationship between confidence and performance, the relationship between average confidence and performance *across individuals* shrinks to approximately zero. For a review of the literature on confidence judgments the reader is referred to Johnson (1955).

A limitation of the usual ordinal scale is that whereas with it one can demonstrate a monotonic relationship between confidence and performance, one cannot conclude whether the degree of confidence is realistic in the sense of being appropriate to the performance or not, either in general or at any specific point on the confidence scale. There are two exceptions to this statement. If the lowest point on the confidence scale indicates that the individual expects his performance at that degree of confidence to be no better than chance, or if the highest point indicates that he expects his performance at that degree of confidence to be perfect, it is then possible to assess the appropriateness at one or both of these points. The finding by Peirce and Jastrow of a discrepancy between expectation and performance at the point of pure guess has been replicated many times since (Adams, 1957a), though the interpretation of this phenomenon is controversial (Goldiamond, 1958).

It seems plausible that the failure to assess realism of confidence accounts for the relative neglect of confidence as an important variable in behavior, despite the reports of studies involving this variable appearing sporadically since 1884. It may be that realism of confidence is in many situations a more important variable than level of performance itself. To take one example, it might be more important, in terms of his future work, social interactions, confidence in himself, etc., for a student to be able to discriminate realistically between what he knows and what he does not know than it would be for him to know considerably more than he does know without such discrimination.[5] Realism of confidence seems ob-

[5] On the basis of some informal observations a hypothesis which seems worth testing is that realism of confidence is a variable which would discriminate between those who have simply "picked up" their general knowledge from conversations and mass communication media and those who have learned through formal education or from authoritative sources.

viously of importance, not only for scholars but in general for individuals making decisions having consequences of importance, either for themselves or for others. To be realistic in the sense of being able to assign probabilities of outcomes accurately does not, of course, imply being "realistic" in the sense of never attempting to accomplish anything which seems highly improbable. The word "cautious" is more accurate for the latter meaning.

PROPOSED METHOD OF MEASURING CONFIDENCE OR SUBJECTIVE PROBABILITY AND REALISM OF CONFIDENCE

The method which we have used is, in a sense, absurdly simple. Like many other psychologists we have done exactly what some writers in the field of subjective probability have asserted to be useless, i.e., we have used direct verbal estimates instead of attempting to infer confidence or subjective probability from other behavior. Those who have shunned this method, while undoubtedly correct with respect to its crude usage in betting situations, have apparently overlooked the frequently demonstrated ordinal relationship between confidence ratings and performance, over a wide range of tasks.[6] We took this relationship as our starting point and simply defined the confidence scale to S in terms of expected percentages (Adams, 1957b), so that S understands that of all those decisions made with confidence p, about $p\%$ should be correct.[7] He understands

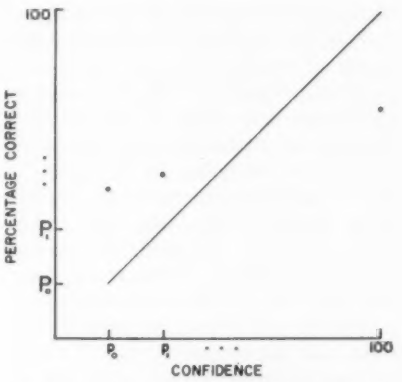


FIG. 1. A realism of confidence function, showing the percentage correct, P, as a function of confidence, p. (The 45-degree line shows the ideal, i.e., $P = p$.)

that if he makes a large number of decisions at each of several points on the confidence scale, then, if he is being realistic in his confidence, the experimental points when plotted should fall near the 45-degree line shown in Figure 1. Definition of the confidence scale in this way has at least two advantages over verbal or numerical ordinal scales: First, performance can be compared with expectation at all points along the scale. Second, individual differences in the meanings of various points on the scale are reduced.

If we say that ideally each decision made with confidence p has independently the probability p of being correct, then the sample percentage correct, P, should be distributed according to the terms of the binomial $(p + q)^n$, in which n is the number of decisions made with confidence p and $q \equiv 1 - p$.[8]

[6] Recent examples of this relationship are provided by L. Decker and I. Pollock (1958) and by E. Carterette and M. Cole (1959). Several other researchers are working along the same lines.

[7] The term "expected percentages" was used because it accurately expresses the intended meaning of the instructions to S. The term "subjective probability" has been used with several different meanings and is not as appropriate.

[8] For theoretical reasons alone we would expect the binomial to give the distribution of P only approximately. First, although it is possible, with some experimental arrangements, to insure that the correctness of the decision of the kth decision is inde-

The expected value of P would therefore be p. Actually P may be much larger or smaller than p; the points plotted in Figure 1 could, if based on sufficiently large sizes of n, show considerable underconfidence at the lower end of the confidence scale ($p < P$) and considerable overconfidence at the upper end ($p > P$). We refer to the extent to which the points P_i approximate the points p_i as *realism of confidence*. Several different realism of confidence scores can be defined; the entire set of deviations $p_i - P_i$, as a function of p_i is called the *realism of confidence function*.

Probably the earliest use of a confidence scale similar to ours was reported by Swets, Tanner, and Birdsall in 1955 in the detection of visual signals. These authors, although referring to the points on their scale as a posteriori probabilities, admittedly treat their data as though the scale were ordinal only. We would go a step further and assert that their data show their observers to be for the most part underconfident in their judgments. The scale under discussion differs from an absolute scale such as weight (in pounds) in that, although S may not have learned to fit his doubts to the scale, the *objective* meaning of the points on the scale is independent of information not immediately available to S, such as the mass of an object kept in the International Bureau of Weights and Measures, or the subjective heaviness of a familiar object such as a pound of margarine. This is not to deny that in some instances making concepts of percentages more concrete by training Ss in the judgment of percentages themselves could conceivably affect subsequent confidence judgments. To the extent that the latter is the case, however, it would also be true that training in judging percentages would modify attitudes toward p values as used in statistical analysis of data.

Let us restate just what the proposed method involves. Ss are asked to attempt to assign a number p to each decision in such a way that the percentage P of correct decisions within a given class (i.e., with a given value of p) will behave like a binomial variate with mathematical expectation p. When we say "behave like a binomial variate" we mean simply that it will be found useful to use the binomial distribution in the analysis of the results which are obtained, and that the set of differences $p_i - P_i$ can be used as a measure of the extent to which S achieved what he was asked to do.[9] Stated in this way, it is clear that this method does not assume that there exists an objectively definable probability that S should ideally give as his confidence judgment for a given decision, i.e., some probability that can be defined independently of his confidence judgment.[10] For example, if S assigns the number 70 to his decision that the statement, "Brussels is the largest city in Belgium," is true, we are not assuming that, given this S's state of knowledge, there is some objectively definable probability that he should give as his confidence judgment; nor are we assuming that there is some objectively

pendent of the correctness of the decision of the $(k-1)$th, it is virtually impossible to prevent some positive or negative autocorrelation of the confidence judgments themselves. Any nonzero autocorrelation will tend to flatten the distribution of P. A second factor working in the same direction is that the confidence judgments at a point p_i actually represent a range of confidence rather than a single degree.

[9] See Footnote 8.

[10] In special cases an objectively definable probability does exist, for example, in guessing experiments in which events are programed by tables of random numbers, or by mechanical random devices. In these cases *each* confidence judgment can be compared with the independently defined probability.

definable probability that Brussels is in fact the largest city in Belgium. Instead, we use the best available information as a criterion for determining the correctness of *S*'s decision, and thus, using the additional decisions made with the same confidence p, we are able to compute P for comparison with p. *S* is instructed that he is not to be concerned with whether the criterion itself (for example, data from the latest census) is invalid. If *S* does not accept the criteria as a basis for the computation of P, he should not accept the results as a measure of realism of confidence. To do so would indicate misunderstanding of the method. Analogous statements can be made for *S*'s decision that a given stock will close at above a given price on a certain future date, etc.

The method we are proposing has three important limitations: (*a*) a suitable criterion for the correctness of each decision must be available, either at the time of decision or later; (*b*) no single confidence judgment itself can be said to be realistic or not (with the exception of certainty paired with an incorrect decision, or zero confidence paired with a correct decision); and (*c*) large numbers of decisions and confidence judgments are necessary for reasonably precise determination of realism of confidence functions.

Even with the limitations mentioned, a realism of confidence function can be obtained under a wide variety of conditions. For example, an *S* can give a confidence judgment in conjunction with a psychophysical decision,[11] a true-false decision, a perception, a motor performance, a recall, a performance in a social situation, a prediction, etc. Further, he can give the confidence judgment either before or after his decision, or both before and after. The instructions may or may not include the word "confidence"; the *S* may instead be asked how likely or probable it is that his decision is correct. The scale extends from $100/K$ to 100, with K being the number of categories among which *S* is to decide.[12] When K is very large, as in tachistoscopic word recognition, the scale extends from 0 to 100.[13] The interpretation of realism of confidence scores depends upon the conditions under which the scores are obtained; the quantitative comparability of the scores under differing conditions, however, offers unusual opportunities for making comparisons and studying interrelationships among psychological processes.

[11] To avoid confusion we shall refer to any operation about which a confidence judgment is made as a "decision."

APPLICATIONS OF THE METHOD

We can illustrate the realism of confidence function with examples from two experimental procedures which differ greatly from each other. Figure 2 shows realism of confidence functions which were obtained from the sorting of 100 true-false statements on each of five consecutive days, a confidence judgment (on a scale from 50 to 100) being obtained with each decision. A different set of statements was used each day; they were of a general nature, being made up from the World Al-

[12] Difficulty of interpretation may arise if the actual events are not chosen by the experimenter in equal numbers from each of the K categories. For example, a response bias toward "true" when *S* is "just guessing" could result in a spurious underconfidence at $p = 50$ if the statements were in fact predominantly true, or a spurious overconfidence if the statements were predominantly false.

[13] It is possible to use the 0 to 100 scale in all cases if *S* is asked to state his confidence that each event falls in a particular category, for example, that each statement is true. We are just beginning to explore results with this method, for comparison with the method previously described.

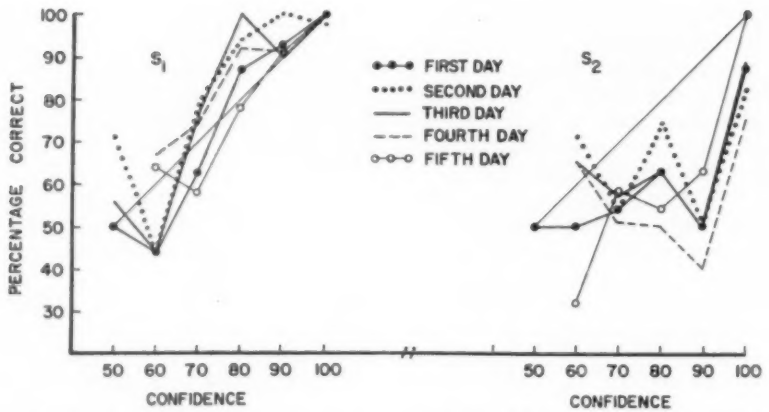


FIG. 2. Realism of confidence functions for 100 true-false decisions about general information statements on each of five consecutive days, for a psychologist and a schizophrenic patient. (Five hundred different statements were used.)

manac and other general information sources and including items from the fields of history, geography, literature, art, sports, and entertainment. The left side of Figure 2 shows the results from S_1 (a psychologist); the right side shows the results from S_2 (a hospitalized schizophrenic). We hasten to add that we are not asserting that the differences between these sets of functions are typical of differences between psychologists and hospitalized schizophrenics. S_1's confidence judgments can be seen to be highly realistic; S_2 can be seen to be highly overconfident, especially at confidence judgments of 80 and 90. This overconfidence is especially striking in view of the fact that S_2 was informed of his overconfidence at the end of each day's session and at the beginning of the next (S_1 was given no information). S_2 acknowledged that his performance demonstrated overconfidence, and made

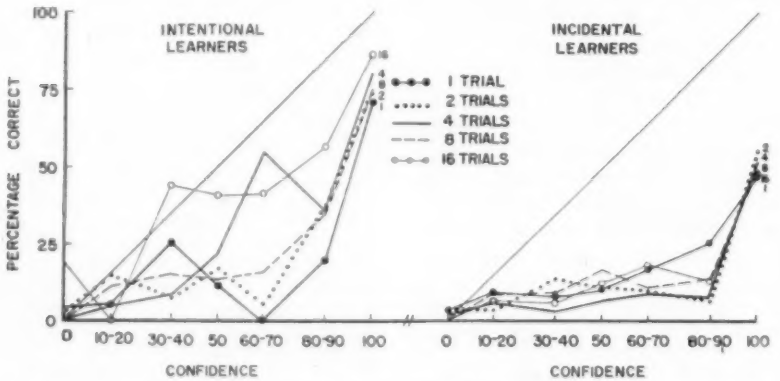


FIG. 3. Realism of confidence functions for recall of nonsense syllables for intentional and incidental learning groups, with 1, 2, 4, 8, or 16 trials.

a strong effort to correct it, but his only improvement lay in achieving, on the fifth day, 100% correct at confidence of 100. S_1's percentage correct increases appropriately as confidence increases whereas S_2's percentage correct in the range 50–90 has a flat appearance, indicating either that S_2 was not differentiating his feelings of doubt and assigning numbers accordingly or that his differential doubts within this range had no validity, at least for these general information items. S_2's dogmatic manner in making assertions was in harmony with his experimental results. He did not, however, question the criteria. We feel justified in labeling this patient "schizophrenic" on the basis of a number of sound criteria, one being his assertion that at times he was Jesus Christ.

Figure 3 shows mean realism of confidence functions for university undergraduates who were asked to indicate their confidence in their recalls of nonsense syllables after 1, 2, 4, 8, or 16 trials on a list of 30 syllables (20 *S*s in each group). The curves on the left are for intentional learners; those on the right are for incidental learners. Both groups show marked overconfidence,[14] but the incidental learners show a greater overconfidence than the intentional learners. Further, the intentional learners give more realistic confidence judgments the more practice they have had on the syllables, whereas the incidental learners do not improve in their realism of confidence. These results indicate that at least in some situations, along with the usual learning process, i.e., an improvement in performance with practice, there occurs an improvement in the ability to assess one's performance in an appropriate or realistic manner.

In general, applications of the method fall into three areas:

1. Generality of realism of confidence, within or between individuals
2. Training in realism of confidence and transfer to different classes of conditions
3. Relations between realism of confidence and other variables, either within or between individuals

In Problem Area 1 we have merely scratched the surface, and shall confine our remarks to a description of this area and what we believe to be its significance. Ordinarily the possibilities of studying the generality of any psychological process within the individual are severely limited by changes in the unit of measurement from one set of conditions to another. For example, it is very difficult to compare learning of nonsense syllables with learning of history *within the individual,* because no comparable units of measurement can be defined. The same can be said for comparing perception of words with perception of colors, etc. Comparisons between psychological processes of this kind are ordinarily possible only in terms of individual differences. Clusters within individuals, however, can never be determined by the study of individual differences, if different individuals have different clusters. Because realism of confidence is measured in the same units regardless of the kind of decision to which it pertains, it is possible to compare realism of confidence for decisions involving different sense modalities, different psychological processes (for example, perception, mem-

[14] The overall measure of overconfidence for each *S* is the *algebraic discrepancy score,* defined as

$$\frac{\Sigma(p_i - P_i)n_i}{\Sigma n_i}$$

in which P_i is the percentage correct at confidence p_i and n_i is the number of decisions made with confidence p_i. This score reduces simply to the mean confidence minus the total percent correct.

ory, reasoning), and different motivational variables such as stress, risk, and anxiety. Presumably "clusters" of realism of confidence functions could be found within the individual which would be interpretable within his overall behavior. So far we have been unable to formulate meaningful hypotheses about clusters of realism of confidence functions within individuals, nor have we yet determined any such clusters experimentally. Of interest also is the question of generality in terms of individual differences, which we have demonstrated among several tasks but have not yet succeeded in organizing in any meaningful way.

Most of our research has been in the second area, i.e., in training *S*s to make more realistic confidence judgments and in the transfer of this training to confidence judgments about radically different decisions. In one experiment, reported in detail elsewhere (Adams & Adams, 1958), we presented *S* with a sequence of 156 pairs of words on each of 5 days, using a different set of pairs each day. *S* decided whether the second word was synonymous with, antonymous to, or unrelated in meaning to the first, then gave his confidence that his decision was correct. The confidence scale ran 33, 40, 50, . . . , 100. At the end of each of the first four sessions each *S* in the experimental group was shown the realism of confidence function that he had produced by his previous 156 decisions and confidence judgments. The control *S*s were shown only their distributions of confidence judgments. The experimental group of 14 *S*s decreased their mean absolute discrepancy score[15] from 13.20 to 6.92 ($p < .01$), whereas the control group *increased* from 11.16 to 15.22 (not significant). This result is interesting when we consider the question of what cue the *S*s learned to respond to. It seems plausible that this cue was an internal one, an event aroused by the external stimulus, the pair of words. Processes of differentiation and discrimination of these subtle events seemed to be occurring.[16] We conceptualize these processes as of three kinds: first, a differentiation of inner events so that their occurrences distribute themselves more readily into a greater number of discrete categories; second, a greater and more discriminating awareness of these events (which are represented in awareness as feelings or attitudes of doubt); and third, a more appropriate fitting of the phenomenally available events to the numbers (expected percentages) on the confidence scale. As the words were of different frequencies, however, and as the percentage correct is related to frequency, it was conceivable that the *S*s were learning to differentiate and discriminate familiarity only, i.e., to match familiarity with confidence judgments.

To determine whether a process of differentiation and discrimination of feelings of doubt might occur, a transfer experiment was performed, with the *S*s (Stanford undergraduates) be-

[15] The *mean absolute discrepancy* score is defined for each *S* as

$$\frac{\Sigma |p_i - P_i| \sqrt{n_i}}{\Sigma \sqrt{n_i}}$$

with meanings as given in Footnote 13. The weight $\sqrt{n_i}$ was used because of its relation to reliability; however, substituting n_i for $\sqrt{n_i}$ did not affect the conclusions.

[16] Certain alternative explanations, such as that the *S*s decreased their discrepancy scores simply by raising or lowering all their confidence judgments, are discussed and rejected by Adams and Adams (1958). The extent to which increased motivation accounts for these results, as well as those of the transfer experiment which followed it, is indeterminate. Our statements about the processes involved in improvement leave open the question of the relative roles of motivation and of practice and feedback per se.

ing trained on decisions different in kind from those on which they were finally tested.[17] On the first day *S* made 108 decisions about the percentage of blue dots in a set of blue and red dots and the percentage of total length a blue line was of a blue and red line. On the second and fourth days *S* decided on the truth or falsity of 250 general statements. On the third day *S* lifted weights while blindfolded and decided upon the ordinal position of the weight in the series. Confidence judgments were given with all decisions, with the scale varying from 10–100 for the percentage decisions to 50–100 for the true-false decisions. Each of the eight experimental *S*s was shown his realism of confidence function at various points in the training; the eight control *S*s were shown only their distributions of judgments, as before. On the fifth day each *S* made 256 decisions as to the synonymity, antonymity, or unrelatedness of word pairs, as in the previous experiment, giving confidence judgments ranging from 33 to 100. All *S*s were run individually. The mean absolute discrepancy score for the experimental group was 9.28, whereas that for the control group was 13.61 ($t = 2.79$; $p < .02$). Comparing these results with those obtained in the previous experiment, we conclude that some transfer of training did occur, though as we would expect it was not 100% transfer.

We believe these results are of theoretical interest for several reasons. First, the explanatory mechanism that seems most plausible to us is some kind of differentiation and discrimination of internal cues—"feelings of doubt"—aroused by the demands of an external situation. Second, the particular content of this learning process may have far reaching implications. In a sense, our *S*s became more realistic in their expectations about the external world, as well as about their own performances. This method of training may provide a way of improving the ability to discriminate knowledge from faith and opinion. There may be important applications of this kind of training, for example, in formal education, although our one attempt to use this method in a classroom situation failed to show positive results; an attempt to train *S*s in groups in a laboratory experiment was also unsuccessful. Of course it would also be possible to train *S*s to be unrealistic—to be overconfident or underconfident. Training of the latter kind would be reminiscent of level-of-aspiration experimentation (Lewin, Dembo, Festinger, & Sears, 1944), in which apparently no *S* has been deliberately trained to be more realistic.

Although this method of training does not solve the classical epistemological problem of finding objective criteria for beliefs, it may be useful as a basis for the development of the ability to weigh evidence soundly. It can be argued that realism of confidence over a specified class of assertions with known truth values is a necessary, though not a sufficient, condition for realism over a related class of assertions with unknown truth values. In other words, there may be little reason to attach much weight to an individual's strong convictions about a given assertion with unknown truth value if there is demonstrably little reason to attach much weight to that same individual's strong convictions about a related class of assertions with known truth values. By "related" we mean to imply some kind of independently determined comparability in the decision and confidence judgment proc-

[17] David Reynolds ran the *S*s in these experiments.

esses, from one class of assertions to another.

The third area, that of relations between realism of confidence and other variables, is also one which we have explored only to a very limited extent. Stanford students taking an objective final examination in a one-quarter course in elementary psychology indicated their confidence in each answer. Fifty-six of these *S*s had been studied intensively by Richard Alpert and Ralph Haber [18] during the previous months as to their achievement motivation and achievement anxiety, and several measures of achievement motivation and anxiety and fear of failure on the examination were taken immediately before and after the examination. Correlations of about .36 ($p < .01$) were obtained between absolute discrepancy scores and fear of failure. Overconfidence or underconfidence, as measured by the algebraic discrepancy score, correlated approximately .06 with these measures. In other words, it was not overconfidence or underconfidence per se, but an inappropriate assessment of one's knowledge which was related to fear of failure on this examination. That this relationship was not simply a function of knowledge per se is shown by the correlations of approximately −.16 between total percentage correct and fear of failure.

Each of these areas of research which we have briefly described is vast. Unfortunately, large numbers of decisions and confidence judgments are necessary to obtain reliable measures of realism of confidence. For example, although in principle the application of this method to the study of paranoid systems (or other cognitive systems) is fairly obvious, the preparation of suitable stimulus materials alone, in order to study a given individual's systematic distortions, may be found to be both difficult and extremely time consuming. We hope that additional researchers will be attracted by this approach to some of these problem areas.

[18] Grateful acknowledgment is made to Haber and Alpert for making their data available to us, to Paul Farnsworth for his cooperation in obtaining our data, and to the National Science Foundation Research Grant NSF G-3045 for 2 hours of computer time on the IBM 650.

REFERENCES

ADAMS, J. K. Behavior without awareness. *Psychol. Bull.*, 1957, **54**, 383–405. (a)

ADAMS, J. K. A confidence scale defined in terms of expected percentages. *Amer. J. Psychol.*, 1957, **70**, 432–436. (b)

ADAMS, P. A., & ADAMS, J. K. Training in confidence-judgments. *Amer. J. Psychol.*, 1958, **71**, 747–751.

CARNAP, R. *Logical foundation of probability.* Chicago: Univer. Chicago Press, 1950.

CARTERETTE, E., & COLE, M. A comparison of the receiver operating characteristics for messages received by ear and by eye. *USN ONR tech. Rep.*, 1959, No. 2. (Contract Nonr 233-58, University of California, Los Angeles)

DAVIDSON, D., & SUPPES, P. *Decision making.* Stanford: Stanford Univer. Press, 1957.

DECKER, L., & POLLACK, I. Confidence ratings and message reception for filtered speech. *J. Acoust. Soc. Amer.*, 1958, **30**, 432–434.

FREUD, S. Repression. In, *Collected Papers.* Vol. IV. London: Hogarth, 1925. Pp. 84–97.

GEORGESCU-ROEGEN, N. The nature of expectation and uncertainty. In M. J. Bowman (Ed.), *Expectations, uncertainty, and business behavior.* New York: Social Science Research Council, 1958.

GOLDIAMOND, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, **55**, 373–411.

HULL, C. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

JOHNSON, D. M. *The psychology of thought and judgment.* New York: Harper, 1955.

LEWIN, K., DEMBO, T., FESTINGER, L., & SEARS, P. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the*

behavior disorders. New York: Ronald, 1944. Pp. 333–378.

PEIRCE, C. S., & JASTROW, J. On small differences of sensation. *Mem. Nat. Acad. Sci.*, 1884, **3**, 73–83.

SAVAGE, L. J. *The foundations of statistics.* New York: Wiley, 1954.

SWETS, J., TANNER, W., & BIRDSALL, T. *The evidence for a decision-making theory of visual detection.* (Electronics Defense Group Technical Report No. 40) Ann Arbor: Engineering Research Institute, Univer. Mich., 1955.

VON NEUMANN, J., & MORGENSTERN, O. *Theory of games and economic behavior.* Princeton: Princeton Univer. Press, 1947.

(Received August 20, 1959)

Psychological Review
1961, Vol. 68, No. 1, 46–59

APPLICATION OF STIMULUS SAMPLING THEORY TO SITUATIONS INVOLVING SOCIAL PRESSURE[1]

PATRICK SUPPES AND FRANKLIN KRASNE
Stanford University

Much contemporary social psychological theory appears to be theory of group behavior qua group. While group oriented concepts may be transiently useful in classifying and systematizing the vast experimental literature of social psychology, we hold that ultimately group behavior can be explained entirely in terms of the behavior of the individuals who constitute the group. In other words we regard the theory of social behavior as a highly important special case of the general theory of individual behavior. Furthermore, to our minds recent quantitative formulations of stimulus-response-reinforcement theory provide excellent conceptual tools for effecting such a subsumption.

This is not to claim that we are prepared to give a detailed stimulus-response-reinforcement analysis of every experimental situation now considered important by social psychologists, but it is to claim that the general lines of such an analysis are clear and in fact may be given in detail, as we shall see for a representative example, for a rather large class of social interaction situations. We conceive of an interaction situation as one in which each member of a group (potentially) provides stimuli and reinforcements for every other member of the group with the behavior of each member entirely explicable on an individual basis given the sequence of stimuli and reinforcements impinging on him. Examples of small group experiments analyzed from this point of view are Hays and Bush (1954), Atkinson and Suppes (1958, 1959), Burke (1959), and Suppes and Atkinson (1960).

Here we wish to consider the theory of social comparison processes, an area of social psychology usually spoken of in terms of frames of reference, social pressure, group norms, and the like, whose experimental work is both especially provocative and particularly well suited to precise experimentation.

Using strictly the notions of stimulus, response, and reinforcement, it is natural to construe the kinds of social situations with which we are concerned as classical or "almost classical" discrimination experiments. The problem, of course, is to identify in each experiment just what is to be considered as stimulus, what as response, and especially difficult what as reinforcement. Once the identifications have been made it is easy enough to assume that a response reinforced in some stimulus situation will have an increased tendency to be emitted on future occasions of that situation.

In the social situation objective stimuli and the behavior of other members of a group combine to form the relevant stimulus situations for each subject. Response classes are in principle arbitrary, and if an experiment is well structured for the subject, the experimenter will be easily led to a "natural" classification of responses.

[1] This research was supported by the Rockefeller Foundation and the Group Psychology Branch of the Office of Naval Research. This paper will be included in the Reprint Series of the Stanford Institute for Mathematical Studies in the Social Sciences.

Considerable difficulty arises in making reinforcement identifications, because subjects bring with them to an experiment a large number of covert verbal responses having secondary reinforcing properties. Consequently, we are simply forced to limit the number of reinforcers which we wish to recognize as important and for analytical purposes to ignore the rest. We shall suppose first that social support per se is reinforcing; although admittedly there will frequently be factors working against it, we will consider them explicitly when they seem important. The proprioceptive stimuli produced by an overlearned response have consistently preceded reinforcement, and therefore the making of an overlearned response is self-reinforcing via the mechanism of secondary reinforcement; thus, we shall assume secondly that there is some automatic (secondary) reinforcement for making responses which have been well overlearned in everyday experience. Lastly, there may be experimenter-controlled rewards of a less ambiguous nature such as money payoffs, prizes, and primary drive reduction.

As a specific example of the type of analysis which we are suggesting we shall study in detail in the remainder of this paper an experimental situation which is similar in certain respects to the classic experiments of Sherif (1935) and in other respects to those of Asch (1956). The theoretical ideas which have been described will be embedded in a stochastic model of the general type described by Suppes and Atkinson (1960), which is a variant of the stimulus sampling theory of Estes and Burke.

In the experiment to be described subjects were required to make a choice on the basis of an objective but slightly ambiguous stimulus situation; in particular they were asked to indicate which of two lines they thought was longer. This choice was followed by an indication of what the subjects were instructed to believe was the correct answer.

From a social psychological point of view the subject plus the experimenter form a dyad which is the simplest case of a small group. According to the social orientation the experimenter exerts (social) "pressure" on the subject to modify his choices in the direction which the experimenter calls correct. In this sense there are certain similarities to Asch's (1956) study; however, since the subject does not find out what the experimenter considers as correct until after his own response has been made, one might suppose, following Sherif (1935), that the effect of the social pressure imposed by the experimenter is to modify the subject's frame of reference on future stimulus presentations.

Our approach to the problem will be to treat the experimental situation as a stimulus discrimination experiment. The situation is complicated slightly in that subjects have strongly overlearned a relevant visual discrimination prior to this experiment; we acknowledge the strong effect of this past learning by positing a secondary reinforcer which may operate to accentuate or attenuate the effect of the experimenter-controlled reinforcer of social support.

It is the conflict between past learning and the experimenter-controlled reinforcer which generates social pressure that does not exist in the usual discrimination experiment. The subject holds Opinion A (here interpreted as a response probability) and another person (in this case the experimenter), who has for one reason or another the power to reinforce the subject, holds B, a potentially different opinion. That the person giving the second re-

sponse is the experimenter is irrelevant. It could equally well have been another subject, and by the same token we could easily pass from the dyad to a larger group. This would involve a more laborious but strictly analogous treatment.

THE EXPERIMENT AND ITS THEORY

On each of a sequence of trials, a pair of lines, one slightly longer than the other, was projected for a few seconds on a screen. The subjects were asked to record on answer sheets which of the two lines (labeled Line 1 and Line 2) they thought was longer. They were then told (what they had been instructed to believe would be) the correct answer; however, it was in reality correct only on a randomly chosen subset of the trials.

In order to describe the situation more precisely, let us introduce some notation:

S_1 = Event of projecting a line pair of which Line 1 is longer
S_2 = Event of projecting a line pair of which Line 2 is longer
A_1 = Response of subject on answer sheet indicating that he thinks Line 1 is longer
A_2 = Response of subject on answer sheet indicating that he thinks Line 2 is longer
E_1 = Reinforcing event of experimenter saying, "Line 1 is longer."
E_2 = Reinforcing event of experimenter saying, "Line 2 is longer."
Ω_1 = Secondary reinforcement of A_1 response
Ω_2 = Secondary reinforcement of A_2 response
$\pi_1 = P(E_1|S_1)$
$\pi_2 = P(E_2|S_2)$
$\gamma = P(S_1)$
δ = Probability that a subject makes a correct discrimination in a control experiment where no information is being given him regarding the correctness of his responses

We will be primarily concerned with the behavior of various conditional probabilities of response as a function of π_1, π_2, and δ.

On any trial first a stimulus event (S_1 or S_2), then a response (A_1 or A_2), and then a reinforcing event (E_1 or E_2) occurs. In addition we introduce automatic secondary reinforcing events, Ω_1 and Ω_2, which reinforce the "perceptually correct" response. Thus, on any trial of an experiment the sequence of events may be indicated by the string of symbols,

$$C \rightarrow S \rightarrow A \rightarrow \Omega \rightarrow E \rightarrow C'$$

where C and C' represent conditioning before and after the trial and where we say that a response is *conditioned* to a stimulus event if the response is elicited as a result of that stimulus event. At all times exactly one response is conditioned to a particular stimulus event.

Independently of what response was actually made on a trial we will say that both E_k and Ω_k *reinforce* A_k. If a reinforcement is *effective*, then the reinforced response becomes conditioned to the stimulus event occurring on the trial. If no reinforcement is effective, then conditioning remains unchanged. It will be assumed that exactly one of the following occurs on every trial: social support is effective, secondary self-reinforcement is effective, neither is effective. Since these events are mutually exclusive and exhaustive, we define

$$\theta_1 = P\ (A_k \text{ is effectively reinforced by } \Omega_k|\Omega_k,\ E_j)$$

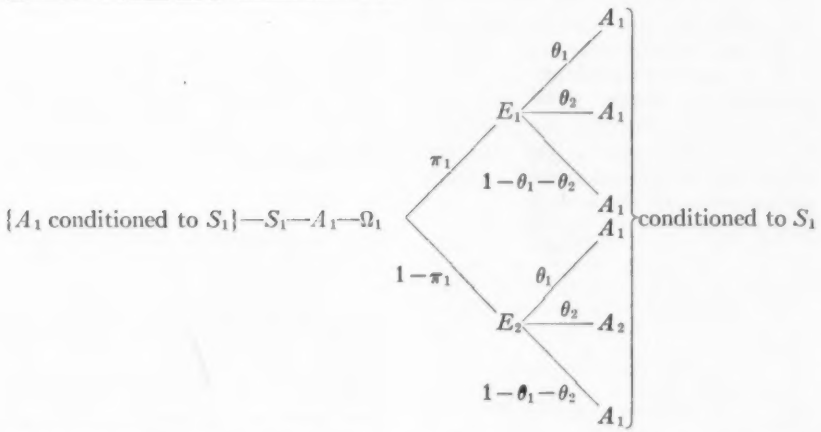
$$\theta_2 = P\ (A_j \text{ is effectively reinforced by } E_j|\Omega_k,\ E_j)$$

$$1 - \theta_1 - \theta_2 = P\ (\text{No reinforcement is effective } |\Omega_k,\ E_j)$$

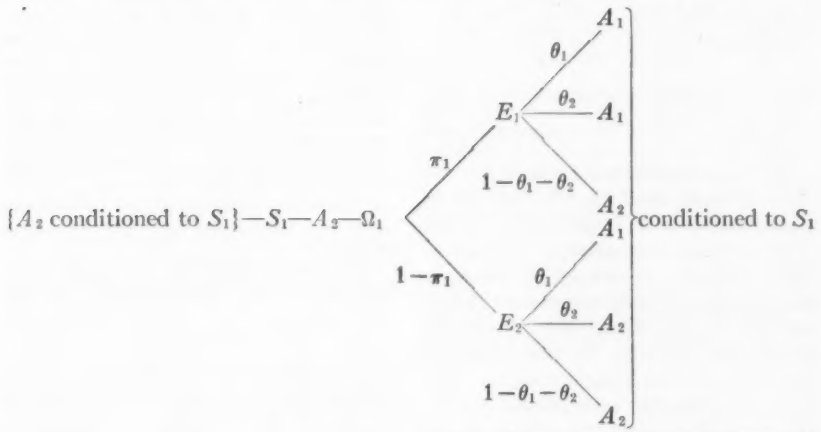
Consider the subsequence of trials on which S_1 and S_2, respectively, occur. Since the conditioning of a response to a stimulus event can be affected only if that stimulus event occurs, we can treat these two subsequences as separate and independent simple learning situations (simple learning as opposed to discrimination

learning). In particular it is easily seen that each learning process is characterized under our assumptions as a Markov chain with A_1 and A_2 as its states (that is to say on any trial either A_1 or A_2 is conditioned to S_i).

As an example of the method used in deriving the transition matrices for these chains, we shall analyze in terms of a tree of logical possibilities a trial on which S_1 occurred. If the subject ended the last S_1 trial conditioned to A_1, then the relevant tree is



and if the subject ended the last S_1 trial with A_2 conditioned to S_1 we have



Using similar methods we arrive at the following Markov chains for the S_1 and S_2 learning processes, respectively. On S_1 type trials we identify two states, 1 and 2, referring to A_1 and A_2, respectively, as being conditioned to S_1; this gives the transition matrix

$$\begin{bmatrix} 1-(1-\pi_1)\theta_2 & (1-\pi_1)\theta_2 \\ \theta_1+\pi_1\theta_2 & 1-\theta_1-\pi_1\theta_2 \end{bmatrix} \quad [1]$$

Similarly on S_2 trials identifying states 1 and 2 referring to A_1 and A_2, respectively, being conditioned to S_2 the transition matrix is

$$\begin{bmatrix} 1-\theta_1-\pi_2\theta_2 & \theta_1+\pi_2\theta_2 \\ (1-\pi_2)\theta_2 & 1-(1-\pi_2)\theta_2 \end{bmatrix} \quad [2]$$

Using the fact that the sequence of S_1's and S_2's occur in accordance with a binomial distribution with parameter γ, we may combine the above processes to obtain $P_n(A_1|S_1)$ and $P_n(A_1|S_2)$ representing the probabilities of A_1 responses on the nth trial of the full experiment given S_1 and S_2 stimulus events, respectively;

$$P_n(A_1|S_1) = \frac{\frac{\theta_1}{\theta_2}+\pi_1}{\frac{\theta_1}{\theta_2}+1} + \left[\delta - \frac{\frac{\theta_1}{\theta_2}+\pi_1}{\frac{\theta_1}{\theta_2}+1}\right] \times [1-\gamma(\theta_1+\theta_2)]^{n-1} \quad [3]$$

$$P_n(A_1|S_2) = \frac{1-\pi_2}{\frac{\theta_1}{\theta_1}+1} + \left[1-\delta - \frac{1-\pi_2}{\frac{\theta_1}{\theta_2}+1}\right] \times [1-(1-\gamma)(\theta_1+\theta_2)]^{n-1} \quad [4]$$

$$n = 1, 2, \cdots$$

The psychophysical constants δ and $(1-\delta)$ have been taken as the initial probabilities of A_1 responses on S_1 and S_2 type trials, respectively. It is clear that since $0 \leq \theta_1 + \theta_2 \leq 1$, the right hand terms of each expression vanish for large n, and hence, the asymptotic results are given by the first terms to the right of the equality signs. We have graphed the asymptotic quantities as a function of the ratio θ_1/θ_2 for relevant values of π_1 and π_2 in Figure 1.

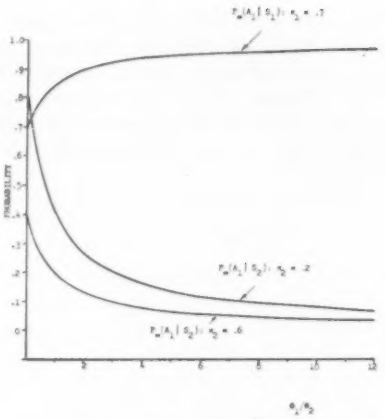


FIG. 1. Asymptotic behavior of $P(A_1|S_1)$ and $P(A_1|S_2)$ as a function of the ratio θ_1/θ_2.

METHOD

Stimuli

The stimuli for these experiments were pairs of parallel lines projected on a field 46 inches wide by 32 inches high; the lines were oriented so that a line passing through their midpoints would be perpendicular to each. Subjects were run under two different stimulus conditions, which we shall designate the "hard" and "easy" conditions to refer to their involving hard and easy discriminations, respectively. Within each difficulty condition (hard or easy) there were two types of slides used (A and B in Table 1), and the orientation of the line pairs was permuted so as to produce $2 \times 2 \times 2 = 8$ different line pairs (e.g., lines vertical, lines horizontal, long line on left, long line on right, etc.). The lines were labeled 1 and 2, respectively, for identification by the subjects. In order to avoid the effect of position habits on the data, for half the subjects the upper horizontal and left vertical lines were always labeled 1 and for the other half the bottom and right lines were called 1.

In Table 1, which gives a description of the line pairs, a is the distance between the lines

TABLE 1

DESCRIPTION OF STIMULI

Slide Type	Hard				Easy			
	a	b	c	δ	a	b	c	δ
A	15.5	5.12	.240	.72	15.5	2.56	.310	.92
B	15.5	7.75	.250	.72	15.5	7.75	.340	.92

Note.—a, b, and c are measured in inches.

TABLE 2

GROUP DESCRIPTIONS

Group Number	π_1	π_2	δ	Description
I	.7	.2	.92	Easy, low symmetry
II	.7	.2	.72	Hard, low symmetry
III	.7	.6	.72	Hard, medium symmetry

of the pair; b is the half-length of the shorter line; c is the difference in half-lengths between the longer and shorter lines; δ is the probability of the subject making a correct discrimination when no experimenter controlled reinforcing events are influencing his behavior.

Room and Apparatus

The line pairs were projected by an Argus 300 Projector with a Sylvania 300-watt projector lamp on to a beaded screen 146 inches away. The subjects (from one to five in number) were seated at a distance of about 96 inches from the front of the screen. The room illumination was about .3 Weston II units.

The subjects responded on answer sheets prepared in essentially the same way as standard IBM score sheets. For half the subjects the left column was used to indicate an A_1 response and for half the right.

Experimental Groups

Three groups of subjects whose conditions are described in Table 2 were run. The degree of symmetry under "Description" refers to relative similarity of reinforcing situations on S_1 and S_2 type trials. For all groups $\gamma = .5$.

Subjects

The subjects, who were students at Stanford University, were obtained from the student employment service, an introductory psychology class, and a university dormitory. With the exception of a few of those from the psychology class all subjects were male. There were 26, 25, and 18 subjects in Groups I, II, and III, respectively. Because of certain technical difficulties in preparing the stimuli, subjects were not placed randomly in groups.

Procedure

There were from one to five subjects per experimental session. The subjects, having been seated and given answer sheets, were given the following instructions. (A sample slide was projected on the screen throughout the instructions.)

> This is an experiment on judgment of length. I am going to flash a number of slides on the screen each of which has two lines on it; your job in each case will be to decide which of the two lines is the longer and to mark your judgment on the answer sheet in the appropriate column opposite the number of the slide we are on. For example: if on the first slide I showed, the line next to the figure one on the screen was longer, you would mark the answer sheet like this [show marked answer sheet]. Sometimes the pair of lines will be vertical rather than horizontal; then the line here would be line one and the one here line two [point to screen to illustrate what I am saying]. Please mark the answer sheets heavily and completely as though you were filling in an IBM score sheet. I will announce the number of the slide we are on before each judgment and will tell you the correct answer after each judgment. You must make your decisions quickly and record them BEFORE I tell you the answer. Each slide will appear for about 2 seconds. One of the two lines will always be longer than the other; however, some of the judgments may be difficult. If you are not sure which line is longer, then guess. Since I want you to work completely independently of each other, I must request that you remain absolutely silent during the experiment. Are there any questions?

The trials then proceeded at a rate of about seven per minute. Each stimulus presentation lasted for approximately 2 seconds. Occasionally the experimenter found it necessary to repeat early in the sessions "this is a test of sorts; you must remain completely silent." Each subject received a total of 250 trials.

After the experiment the subjects were instructed: "Will you please write a brief statement of your reactions to this experiment?" The responses to this question indicated that the subjects understood and believed the instructions.

ESTIMATION METHODS

Two methods will be presented for estimating the parameters θ_1 and θ_2 from the data of our experiments. First we shall make maximum likelihood estimates based on the separate S_1 and S_2 response subsequences

(Anderson & Goodman, 1957); then, taking advantage of the mutual independence of the S_1 and S_2 learning processes, we shall maximize the joint likelihood function to obtain single estimates of θ_1 and θ_2.

It can be shown that the first method is essentially equivalent to setting the expressions for the theoretical transition probabilities equal to their estimates and solving for θ_1 and θ_2.

The estimates based on maximizing the joint likelihood function are not quite so simple. The derivation, which is reasonably straightforward but tedious, leads to a tenth degree equation for the unknown, θ_2; solutions were obtained by approximation methods.

RESULTS AND DISCUSSION

Tables 3 and 4 present learning parameter estimates obtained by the two methods described above.

The most important observation to be made is that for all estimates θ_1/θ_2 is less on hard than on easy discriminations. One can go even further and point out that with the exception of the S_2 trial estimate in Group III this learning parameter ratio is greater than one for the easy discriminations and less than one for the hard ones. This says essentially (for our experimental situation) that if the physical stimulus situation is fairly unambiguous, subjects will "prefer" to derive their reinforcement from it via Ω_1 and Ω_2 while if it is not, they will depend primarily on social support. This makes good intuitive sense and in fact has been adopted as a postulate by Festinger (1954) in his theory of social comparison processes.

In spite of the fact that we get these intergroup differences for all estimates, it is interesting and rather odd that, as will be seen by studying

TABLE 3

SEPARATE MAXIMUM LIKELIHOOD ESTIMATES OF θ_1, θ_2, AND DERIVED QUANTITIES

Type Trials on Which Estimates Based	Quantity	Group I	Group II	Group III
S_1	$\hat{\theta}_1$	.60	.38	.30
	$\hat{\theta}_2$	.20	.54	.62
	$\hat{\theta}_1/\hat{\theta}_2$	3.0	.71	.48
	$\hat{\theta}_1 + \hat{\theta}_2$	.78	.93	.91
S_2	$\hat{\theta}_1$	.42	.36	.48
	$\hat{\theta}_2$	.27	.36	.45
	$\hat{\theta}_1/\hat{\theta}_2$	1.6	1.0	1.1
	$\hat{\theta}_1 + \hat{\theta}_2$	.69	.72	.92

Table 3, the effects described above show up much more markedly on S_1 than on S_2 type trials. Since the actual stimulus and response labeling was randomized over subjects, and since the stimuli and responses were physically symmetrical, the relation between π_1 and π_2 must be responsible for this rather surprising phenomenon.

Reference to the tables under discussion will show that the over-all effectiveness of reinforcement, $\theta_1 + \theta_2$, increased from the easy to the hard conditions. One might argue that the harder a discrimination, the more attentive the subject will be to all events surrounding him; thus, reinforcements are generally more effective when δ is small. Although it is true that via Equations 3 and 4 one can estimate the ratio θ_1/θ_2 directly from asymptotic data, we have preferred to make much more efficient

TABLE 4

JOINT MAXIMUM LIKELIHOOD ESTIMATES OF θ_1, θ_2, AND DERIVED QUANTITIES

Quantity	Group I	Group II	Group III
$\hat{\theta}_1$	.44	.39	.40
$\hat{\theta}_2$	.25	.42	.52
$\hat{\theta}_1/\hat{\theta}_2$	1.8	.92	.77
$\hat{\theta}_1 + \hat{\theta}_2$	.69	.81	.92

estimates of this ratio by making strong use of the Markov properties of response subsequences. Specifically, the ratio $\hat{\theta}_1/\hat{\theta}_2$, where $\hat{\theta}_1$ and $\hat{\theta}_2$ are maximum likelihood estimates, yields the maximum likelihood estimate of θ_1/θ_2. Consequently, we wish to emphasize that the extent of agreement between observed asymptotic behavior and the asymptotes *predicted* using the maximum likelihood estimates of θ_1 and θ_2 is an empirical question representing one sort of test of the model. The predicted asymptotic conditional response probabilities have been included in Table 5, which also presents for purposes of comparison the relevant relative frequencies over the last 50 and 100 trials for each group. Although the fit is not perfect, the observed and predicted rank orders across groups are in agreement for the predictions based on the joint maximum likelihood estimates.

Observed mean learning curves on which the predicted asymptotes (of the last line of Table 5) are indicated as dashed lines are plotted in Figure 2. In order to determine whether a significant amount of learning occurred, the conditional response proportions from the first and last 50 trials of each experiment were compared by two-tailed sign tests (Siegel, 1956); the significance levels α are shown next to the appropriate curves.

The sign test, which was used here because the distribution assumptions necessary for it are almost negligible, does not produce acceptable significance levels. Higher level ($\alpha < .01$) tests may be obtained if we are willing to make the rather strong assumption that the individual responses of a particular subject are independent Bernoulli events. However, this assumption is too grossly violated to allow one to place much confidence in results based upon it.

If we draw learning curves (Equations 3 and 4) using the maximum likelihood estimates of θ_1 and θ_2, we find that the curves rise to their asymptotes much more rapidly than do the observed curves of Figure 2. We wish to point out that this does not represent a fundamental problem for the present theory although it limits the usefulness of the specific form of the model which we have presented. The problem arises from our having assumed that only one stimulus "element" or "pattern" (see Estes, 1959) was associated with each stimulus event, S_i. By positing a larger number of such stimulus patterns and retaining the supposition that exactly one is sampled on each trial, the

TABLE 5

OBSERVED AND PREDICTED ASYMPTOTIC CONDITIONAL RESPONSE PROBABILITIES

	Group I		Group II		Group III	
	$P(A_1\|S_1)$	$P(A_1\|S_2)$	$P(A_1\|S_1)$	$P(A_1\|S_2)$	$P(A_1\|S_1)$	$P(A_1\|S_2)$
Observed: Over last 100 trials	.92	.31	.83	.40	.79	.19
Observed: Over last 50 trials	.94	.28	.84	.39	.74	.18
Theoretical: Based on S_1 trial estimates	.92	.20	.83	.48	.79	.27
Theoretical: Based on S_2 trial estimates	.88	.31	.85	.40	.85	.19
Theoretical: Based on joint MLE	.89	.29	.84	.42	.83	.23

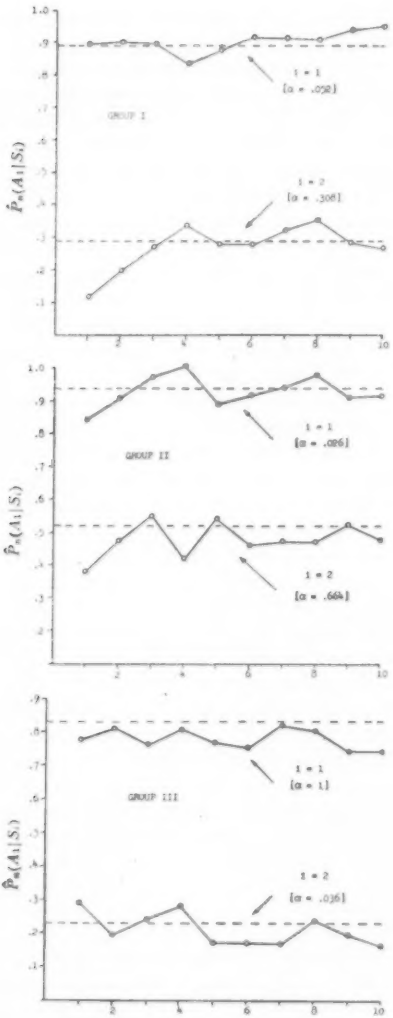


FIG. 2. Mean learning curves. (Blocks of 25 trials.)

theoretical learning rate may be decreased at the expense of considerably more tedious computations.

We may take a compromise approach, however. If we suppose there to be N stimulus elements associated with each stimulus situation, then although most formulae become considerably complicated, the learning curves 3 and 4 are changed only in the last term; let $\theta_1/\theta_2 = \epsilon$ then we may write

$$P_n(A_1|S_1) = \frac{\epsilon+\pi_1}{\epsilon+1} + \left(\delta - \frac{\epsilon+\pi_1}{\epsilon+1}\right) \times \left[1-\gamma\frac{\theta_2}{N}(\epsilon+1)\right]^{n-1}$$

$$P_n(A_1|S_2) = \frac{1-\pi_2}{\epsilon+1} + \left(1-\delta - \frac{1-\pi_2}{\epsilon+1}\right) \times \left[1-(1-\gamma)\frac{\theta_2}{N}(\epsilon+1)\right]^{n-1}$$

We may estimate ϵ directly from asymptotic data and choose θ_2/N to make a least squares fit of the theoretical to the empirical learning curves.

CONCLUSIONS

In conclusion it is our opinion that considerable analytic advantage has derived from application of a quantitative theory of individual behavior in the present context. Through its use we have been able to make highly specific remarks about the extent of operation of certain events which are not directly observable but which may reasonably be postulated to exist, namely secondary self-reinforcement of responses well established in everyday experience. Also we have been able to compare the magnitude of the effects of these secondary reinforcements to that of the more directly controlled social support. As far as we can see, it would be considerably more difficult to obtain comparable information from experimentation not oriented about a quantitative theory.

We wish to emphasize again that the central ideas behind the present analysis are ones of considerable generality; similar techniques of analysis can be applied to many social

psychological situations by a fairly straightforward extension of the ideas presented here.

REFERENCES

ANDERSON, T. W., & GOODMAN, L. A. Statistical inference about Markov chains. *Ann. math. Statist.*, 1957, **28**, 89–109.

ASCH, S. E. Studies of independence and submission to group pressure: 1. A minority of one against a unanimous majority. *Psychol. Monogr.*, 1956, **70**(9, Whole No. 416).

ATKINSON, R. C., & SUPPES, P. An analysis of two-person game situations in terms of statistical learning theory. *J. exp. Psychol.*, 1958, **55**, 369–378.

ATKINSON, R. C., & SUPPES, P. Application of a Markov model to two-person noncooperative games. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959. Ch. 3.

BURKE, C. J. Applications of a linear model to two-person interactions. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959. Ch. 9.

BUSH, R. R., & ESTES, W. K. (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959.

ESTES, W. K. Component and pattern models with Markovian interpretations. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959. Ch. 1.

FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, 1954, **7**, 117–140.

HAYS, D. G., & BUSH, R. R. A study of group action. *Amer. sociol. Rev.*, 1954, **19**, 693–701.

SHERIF, M. A study in some social factors in perception. *Arch. Psychol.*, 1935, No. 187.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

SUPPES, P., & ATKINSON, R. C. *Markov learning models for multiperson interactions.* Stanford: Stanford Univer. Press, 1960.

(Received October 5, 1959)

APPENDIX A

The model (Model I) originally proposed for the experimental situation studied in the body of this paper differed in its reinforcement mechanisms from the one (Model II) actually discussed there. Since rather extensive work was done on it and since we consider some of its failings to be instructive, we present here some detailed comments on Model I.

Let us call those trials on which the subscripts of E_i and S_j do not agree "conflict trials." Rather than introducing secondary reinforcers Ω_1 and Ω_2, we assume that the reinforcing event E_k is effective with probability θ_A on nonconflict trials and with probability $\theta_B < \theta_A$ on conflict trials. While we make no explicit psychological assumptions regarding the reasons for attenuating the learning parameter on conflict trials, the intuitive reasons are clear.

Markov process state identifications are exactly the same as in Model II, and we here present those results for Model I which correspond to those indicated for Model II in the body of the paper. On S_1 trials the transition matrix is

$$\begin{bmatrix} 1-(1-\pi_1)\theta_B & (1-\pi_1)\theta_B \\ \pi_1\theta_A & 1-\pi_1\theta_A \end{bmatrix} \quad [A1]$$

and on S_2 trials it is

$$\begin{bmatrix} 1-\pi_2\theta_A & \pi_2\theta_A \\ (1-\pi_2)\theta_B & 1-(1-\pi_2)\theta_B \end{bmatrix} \quad [A2]$$

The learning curves are

$$P_n(A_1|S_1) = \frac{\pi_1}{\pi_1+\frac{\theta_B}{\theta_A}(1-\pi_1)} + \left(\delta - \frac{\pi_1}{\pi_1+\frac{\theta_B}{\theta_A}(1-\pi_1)}\right) \times \{1-\gamma[\theta_B(1-\pi_1)+\theta_A\pi_1]\}^{n-1} \quad [A3]$$

$$P_n(A_1|S_2) = \frac{1-\pi_2}{\frac{\theta_A}{\theta_B}\pi_2+(1-\pi_2)} + \left((1-\delta) - \frac{1-\pi_2}{\frac{\theta_A}{\theta_B}\pi_2+(1-\pi_2)}\right) \times \{1-(1-\gamma)[\theta_B(1-\pi_2)+\theta_A\pi_2]\}^{n-1} \quad [A4]$$

And the graphs of the asymptotes as a function of θ_A/θ_B, which are presented in Figure A1, should be compared with the corresponding ones for Model II.

Several methods of learning parameter estimation were used for this model, and each led to the same rather disquieting result: for all methods the estimates of θ_A were greater than one. Before considering the reason for this anomaly, let us consider the various methods of estimation from which we obtained estimates of θ_A and θ_B.

Method 1. This method is precisely the same in principle as the first method presented in the body of the paper for θ_1 and θ_2. It is a true maximum likelihood estimate on each of the two response subsequences. As in Model II the method is for a fairly large number of counted transitions per subject equivalent to setting the observed transition matrices equal to the theoretical ones and solving. A comparison of the transition matrices for the two models shows that using this estimation procedure $\hat{\theta}_2 = \hat{\theta}_B$.

Method 2. The joint maximum likelihood estimate here, which is considerably simpler than in Model II, may be obtained explicitly by solving a second degree equation.

Method 3. The next procedure, which is rather complicated in the present experiments, we consider especially interesting, because it seems to yield many of the advantages of maximum likelihood estimation in situations where true maximum likelihood estimation is unfeasible.

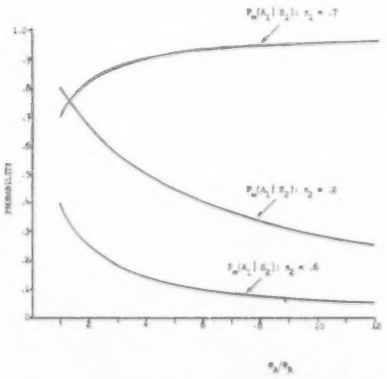


FIG. A1. Asymptotic behavior of $P(A_1|S_1)$ and $P(A_1|S_2)$ as a function of the ratio θ_A/θ_B in Model I.

Consider the full sequence of responses (S_1 and S_2 type trials in the same sequence now) for a particular subject. It is possible to compute for large n

$$\lambda_{ij} = P_\infty\,(A_j \text{ on Trial } n \,|\, A_i \text{ on Trial } n-1)$$

independently of the trial number, n. Thus, while it is true that the sequence of responses for a subject is a chain of infinite order, we may as a first approximation treat it as a stationary Markov chain with transition probabilities, λ_{ij}, and compute under this assumption the maximum likelihood estimate of the parameter pair (θ_A, θ_B); such a procedure is an example of *pseudo maximum likelihood estimation*. The pseudo likelihood function is

$$L(\theta_A, \theta_B) = \prod_{i,j} \lambda_{ij}^{n_{ij}} P_\infty^{n_1}(A_1) P_\infty^{n_2}(A_2)$$

where $P_\infty(A_i)$ is taken asymptotically, n_i is the number of occurrences of response A_i on the first trial of the sequence over which the pseudo maximum likelihood estimate is being taken, and n_{ij} is the observed number of transitions from response A_i to A_j over the sequence in question.

Method 4. One simple estimate of θ_A/θ_B was used in order to provide a check on the above procedures. Let $\hat{P}_\infty(A_1|S_1)$ and $\hat{P}_\infty(A_1|S_2)$ be the estimated (over the last 50 trials) asymptotic conditional probabilities of response. Then the value of θ_A/θ_B which minimizes

$$\left|\hat{P}_\infty(A_1|S_1) - \frac{\pi_1}{\pi_1 + \frac{\theta_B}{\theta_A}(1-\pi_1)}\right| + \left|\hat{P}_\infty(A_1|S_2) - \frac{1-\pi_2}{\frac{\theta_A}{\theta_B}\pi_2 + (1-\pi_2)}\right|$$

is a rather natural estimate of θ_A/θ_B although it uses very little of the information in the data.

The various estimates of θ_A and θ_B are given in Table A1.

TABLE A1

MAXIMUM LIKELIHOOD ESTIMATES OF θ_A, θ_B, AND θ_A/θ_B

Estimation Method	Quantity	Group		
		I	II	III
1	θ_A on S_1 trials	1.0	1.1	1.0
	θ_B on S_1 trials	.20	.54	.62
	θ_A/θ_B on S_1 trials	5.3	2.0	1.7
	θ_A on S_2 trials	2.4	2.2	1.1
	θ_B on S_2 trials	.27	.36	.45
	θ_A/θ_B on S_2 trials	8.9	6.0	2.4
2	θ_A	1.3	1.2	1.1
	θ_B	.24	.42	.52
	θ_A/θ_B	5.2	2.9	2.2
3	θ_A	2.2	1.7	.56
	θ_B	.22	.16	.07
	θ_A/θ_B	10.0	11.0	8.0
4	θ_A/θ_B	10.0	6.0	2.4

The occurrence of learning parameter estimates greater than one makes interpretation of Model I very tenuous. It will be instructive to consider the problem in more detail. It will be seen from Equations A1 and A2 on which our estimation procedures are primarily based, that θ_A and θ_B appear in no equations which would restrict their values to the unit interval. It is only in interpretation and in the derivation of transition matrices that we have made use of the learning parameters' roles as probabilities.

Although we may obtain bounded estimates of θ_A and θ_B by looking at detailed sequential properties of the data, we find that it is certain inadequacies of the learning model at precisely this level of analysis which are the cause of our excessively large learning parameter estimates.

To be more specific let $A_{1,n}$ represent response A_1 on trial n, and let $E_{1,n}$ be defined analogously. Then, it easily follows from our learning assumptions that on S_1 type trials

$$P(A_{1,n+1} | E_{2,n}, A_{2,n}) = 0 \qquad [A5]$$

$$P(A_{1,n+1} | E_{1,n}, A_{2,n}) = \theta_A \qquad [A6]$$

Although it would now seem quite natural to use Equation A6 as the basis for a bounded estimate of θ_A, we find that it is precisely because Equation A5 is not empirically valid that our previous estimation procedures have yielded estimates of θ_A greater than one. It can be easily shown that using estimation Method 1, for example, θ_A is indeed restrained to the unit interval if the observed relative frequency associated with the left side of Equation A5 is near zero, which in the present study it unfortunately is not.

In spite of its other difficulties the present model gives a sufficient description of important qualitative aspects of the data. First, *θ_A is consistently greater than θ_B*; second, *comparing Group I (easy, low symmetry) and Group II (hard, low symmetry), for which π_1 and π_2 are the same, Group I, which involved the easier discrimination, has a greater learning parameter ratio θ_A/θ_B.* Although the pseudo maximum likelihood estimates of the learning parameters weakly contradict this latter statement, it is our feeling that we may place more confidence in the true maximum likelihood estimates, since their optimality properties are well known.

We should expect under this model no change in the θ values as a function of π_1 and π_2 when the same stimuli are used. Unfortunately there is such a change;

indeed, we notice that when the subject makes the less frequently reinforced response and *is* reinforced, this reinforcement tends to have a greater probability of being effective than it would have otherwise had. This can be seen by comparing the θ estimates made on S_1 and S_2 type subsequences; the θ_B estimates in Group I, however, are contrary to this generalization.

APPENDIX B

We consider it to be of some general interest to present the method of obtaining the learning curves 3, 4, 7, and 8, since essentially the same technique may be used to derive learning curves in an N element pattern model (Estes, 1959) with a probability distribution over the elements.

Let

$$\sigma_{i,n,k} = \left\{\begin{matrix}\text{Event of precisely } k \text{ samplings} \\ \text{of stimulus element } i \text{ in first} \\ n \text{ trials, } n = 1, 2, \cdots\end{matrix}\right\}$$

Let

$$\alpha_{n+1}{}^{(i)} = P(A_{1,n} | S_{i,n} \cap \sigma_{i,n-1,m}) \quad [B1]$$

Now,

$$P(A_{1,n} | S_{i,n}) = \sum_{\nu=0}^{n-1} \frac{P(A_{1,n} \cap S_{i,n} \cap \sigma_{i,n-1,\nu})}{P(S_{i,n})} \quad [B2]$$

$$= \frac{1}{P(S_{i,n})} \sum_{\nu=0}^{n-1} P(A_{1,n} | S_{i,n} \cap \sigma_{i,n-1,\nu}) \times P^{\nu}(S_{i,n} | \sigma_{i,n-1,\nu}) P(\sigma_{i,n-1,\nu}) \quad [B3]$$

$$= \sum_{\nu=0}^{n-1} \alpha_{\nu+1}{}^{(i)} \binom{n-1}{\nu} P^{\nu}(S_i) \times (1 - P(S_i))^{n-1-\nu} \quad [B4]$$

In a general 2×2 Markov process if x_n is the probability of being in state one on trial n and

$$P = \begin{pmatrix} 1-b & b \\ a & 1-a \end{pmatrix} \quad [B5]$$

then

$$x_n = (1-a-b)^n \left[x_0 - \frac{a}{a+b} \right] + \frac{a}{a+b} \quad [B6]$$

here, $n = 0, 1, \cdots$.

But for the transition matrices for S_1 and S_2 type trials, respectively, we have

S_1 trials:
$$\begin{aligned} b &= (1 - \pi_1)\theta_2 \\ a &= \theta_1 + \pi_1\theta_2 \end{aligned} \quad [B7]$$

so

$$a + b = \theta_1 + \theta_2$$

S_2 trials:
$$\begin{aligned} b &= \theta_1 + \pi_2\theta_2 \\ a &= (1 - \pi_2)\theta_2 \end{aligned} \quad [B8]$$

so

$$a + b = \theta_1 + \theta_2$$

Thus,

$$\alpha_{\nu+1}{}^{(i)} = \alpha_{\infty}{}^{(i)} + (\alpha_0{}^{(i)} - \alpha_{\infty}{}^{(i)}) \times [1 - \theta_1 - \theta_2]^{\nu} \quad [B9]$$
$$\nu = 0, 1, \cdots,$$

which substituting in Equation B4 gives

$$\begin{aligned} P(A_{1,n} | S_{i,n}) &= \sum_{\nu=0}^{n-1} [\alpha_{\infty}{}^{(i)} + (\alpha_0{}^{(i)} - \alpha_{\infty}{}^{(i)}) \times (1 - \theta_1 - \theta_2)^{\nu}] \binom{n-1}{\nu} \\ &\quad \times P^{\nu}(S_i)(1 - P(S_i))^{n-1-\nu} \quad [B10] \\ &= \alpha_{\infty}{}^{(i)} \sum_{\nu=0}^{n-1} \binom{n-1}{\nu} P^{\nu}(S_i) \times [1 - P(S_i)]^{n-1-\nu} \\ &\quad + (\alpha_0{}^{(i)} - \alpha_{\infty}{}^{(i)}) \sum_{\nu=0}^{n-1} \binom{n-1}{\nu} \times [(1 - \theta_1 - \theta_2)P(S_i)]^{\nu} \times [1 - P(S_i)]^{n-1-\nu} \\ &= \alpha_{\infty}{}^{(i)} + (\alpha_0{}^{(i)} - \alpha_{\infty}{}^{(i)}) \times [(1 - \theta_1 - \theta_2)P(S_i) + 1 - P(S_i)]^{n-1} \\ &= \alpha_{\infty}{}^{(i)} + (\alpha_0{}^{(i)} - \alpha_{\infty}{}^{(i)}) \times [1 - P(S_i)(\theta_1 + \theta_2)]^{n-1} \quad [B11] \end{aligned}$$

This becomes for S_1 trials, letting

$$\alpha_0^{(1)} = \delta$$

$$P(A_{1,n}|S_{1,n}) = \frac{\theta_1 + \pi_1\theta_2}{\theta_1 + \theta_2} + \left[\delta - \frac{\theta_1 + \pi_1\theta_2}{\theta_1 + \theta_2}\right] \times [1 - \gamma(\theta_1 + \theta_2)]^{n-1} \quad [B12]$$

and for S_2 trials becomes, letting

$$\alpha_0^{(2)} = 1 - \delta$$

$$P(A_{1,n}|S_{2,n}) = \frac{(1-\pi_2)\theta_2}{\theta_1+\theta_2} + \left[(1-\delta - \frac{(1-\pi_2)\theta_2}{\theta_1+\theta_2}\right] \times [1-(1-\gamma)(\theta_1+\theta_2)]^{n-1} \quad [B13]$$

Psychological Review
1961, Vol. 68, No. 1, 60–67

AN ANALYSIS OF GSR CONDITIONING

M. A. STEWART, J. A. STERN, G. WINOKUR, AND S. FREDMAN

Department of Psychiatry and Neurology, Washington University School of Medicine, St. Louis

Classical conditioning was defined by a round table conference on nomenclature in animal behavior as

> The process of acquisition by an animal of the capacity to respond to a given stimulus with the reflex reaction proper to another stimulus (the reinforcement) when the two stimuli are applied concurrently for a number of times (Thorpe, 1956).

This definition fits the type of experiment done by Pavlov, in which a specific form of response, namely salivation, is conditioned to a neutral stimulus such as the sound of a metronome. By "specific form of response" in this context we mean a response which is naturally evoked by only a limited class of stimuli. This idea is reflected by Pavlov's own definitions of the terms:

> In the unconditioned reflex, those properties of the substance to which the saliva is physiologically adapted act as the stimulus, for example, the hardness, the dryness, the definite chemical properties, etc.; in the conditioned reflex, on the other hand, those properties which bear no direct relation to the physiological role of the saliva act as stimuli, for example, color, form and the like (Pavlov, 1928).

While theoretically no stimulus can be completely neutral in relation to any response, there is no practical problem in finding stimuli which are neutral in relation to a response such as salivation. But much of the work done in classical conditioning, particularly with human subjects (*S*s), has involved responses which are far from specific in this sense, for instance, GSR, eyeblink, and startle responses. In working with these responses one finds that almost no stimuli are neutral, and hence that the reflex action proper to the conditioned stimulus is the same as that proper to the unconditioned stimulus, though varying in extent. It follows that experimenters in this field should always be aware of the possibility that they are producing "pseudoconditioning," or an augmenting of an original response to the conditioned stimulus rather than true conditioning.

The problem is reviewed by Hilgard and Marquis (1940), who say:

> the augmentation of an original response to the conditioned stimulus through a conditioning procedure may be described as sensitization. This is merely a form of pseudo-conditioning, unless it can be demonstrated that the sensitization is a function of the repetition of conditioned and unconditioned stimuli in precise relationship.

We doubt whether it is right to speak of conditioning rather than sensitization even if this qualification is met. The experiments of Bernstein (1934) and Hunter (1937) showed that in the case of eyeblink responses in man and startle responses in rats, conditioned responses occurred at different latencies than those of the original responses to the conditioned stimuli, and in the case of the rats could further be distinguished from the original responses by the fact that this latency tended to increase with the number of conditioning trials and decrease during extinction trials, while the latency of the original responses remained constant. Our own experience, working with the GSR, has been that true conditioned responses, that is, responses which are not elicited by the

conditioned stimulus before its being paired with the unconditioned stimulus, can be differentiated from other responses to the conditioned stimulus on the basis of latency. We have used this criterion in previous experiments (Stewart, Winokur, Stern, Guze, Pfeiffer, & Hornug, 1959; Winokur, Guze, Stewart, Pfeiffer, Stern, & Hornug, 1959) and further explored its use in the experiment to be described below.

There is no a priori reason for thinking that sensitization does not follow discoverable laws as do habituation and conditioning. Only if this were so would it be reasonable to differentiate sensitization from conditioning on the grounds that the one depended on the repetition of the stimuli in precise relationship and the other did not. In our opinion, therefore, the authors who have paid most attention to this problem previously, White and Schlosberg (1952), were not justified in their assertion that conditioning of the GSR can only be effected by using a .5-second delay period, and that the use of other delay periods results only in pseudoconditioning. We propose to review representative previous papers on conditioning the GSR and show that this work has probably so far dealt only with sensitization, and not with true conditioning.

In the paper referred to above, White and Schlosberg describe an experiment in which they compared the effects of varying the delay period between conditioned and unconditioned stimuli while conditioning the GSR to light. Groups of *S*s were conditioned with delay periods of 0, .25, .5, 1, 2, and 4 seconds and a control group was given the same stimuli unpaired at an interval of 30 seconds, which was one-half the intertrial interval used in the test groups. The authors do not precisely define what responses were called conditioned responses, but say "the primary data consisted of measured deflections on each of the five extinction trials, and on a final shock." No mention is made of the time relationships of these deflections to the stimuli, even though a reference is made to the fact that a response was evoked by the light before conditioning trials began. The number of conditioning trials in each group was five, and no initial habituation trials were carried out. In our experience very few true conditioned responses appear within the first five trials in this type of experiment when normal *S*s are used, while the previously habituated unconditioned response to the conditioned stimulus reappears at the second conditioning trial. We assume, therefore, that the deflections to which the authors refer are most probably augmented unconditioned responses to the conditioned stimulus. The findings of this investigation were that only the group conditioned with the .5-second delay period showed significantly larger conditioned responses than the control group. It is interesting to speculate whether even this significant difference might not have disappeared if the experimenters had used a longer intertrial interval than 30 seconds for the control group. It is well established, for instance by the work of Berg and Beebe-Center (1941), that the adaptation of unconditioned responses to stimuli is accelerated by decreasing the interval between stimuli, and slowed or stopped altogether by lengthening the interval.

Bitterman and Holtzman (1952), in their study of GSR conditioning as a measure of anxiety, say:

> The initial response to tone disappears to a great extent during the unreinforced adaptation trials; during the conditioning trials the tone response grows in negatively accelerated fashion, and then it tends to

decline once more during the extinction series.

The "tone response" is defined later in their paper as "the initial deflection produced by the tone." In a study on GSR conditioning in psychiatric patients, Howe (1958) states:

It was evident in the vast majority of cases that adaptation was then complete, since, at the fourth pretesting presentation of the tone, the response was either zero or negligible.

He does not actually define what he then regarded as a conditioned response to tone, but says that after each stimulation "minimal and maximal galvanometric readings were taken," without specifying any time relations of the responses. Franks (1956) in describing the experimental technique for his study of conditioning in normal and neurotic *S*s, says:

Before the conditioning session was begun, each *S* was given three tone stimuli, followed by three air puff stimuli (not paired with the CS) and then three more tone stimuli. Only *S*s who did not give PGR or eyeblink responses to the last three tone stimuli were included in the conditioning study. The purposes of these trials were to eliminate those *S*s who showed any evidence of pseudo-conditioning or original sensitivity to the tone.

He then later defines an eyelid or GSR conditioned response as a deflection in the record occurring during a latency of between 0.5 and 8.0 seconds after the onset of the CS.

All these authors, then, have left the significance of their findings in doubt for want of definition of the response treated as a conditioned response, and have left ample grounds for the suspicion that their findings do not relate to conditioning but to sensitization of the GSR.

In addition, their methods raise certain general problems of GSR conditioning which invite comment. The first is the soundness of excluding *S*s who give more than a few unconditioned responses to the conditioned stimulus, as was done in Franks' study. This is a purely arbitrary division of *S*s which depends to a large extent on the sensitivity of the recording apparatus. The important result of this exclusion, however, is the skewing of the test group in favor of nonreactive *S*s, which must invalidate any comparison of clinical subgroups such as Franks was interested in. Secondly, the work of Landis and Hunt (1935) and of Cook and Harris (1937) has shown that the threshold for evocation of the GSR and the general reactivity of an *S* measured by the GSR, is markedly affected by uncontrollable factors such as the *S*'s feelings and general factors such as instructions given to the *S* at the outset of the experiment. The implication of this is that the general conditions of the experiment and in particular the introduction to the experiment must be held constant. Finally, it appears from the work of Prosser and Hunter (1936) that a conditioned stimulus which initially evokes no unconditioned response may do so after being paired with a painful unconditioned stimulus a number of times through an alteration of the threshold for evocation of the response. These authors found that when a sound, which initially evoked no startle response, had been presented to rats together with a shock a number of times, the sound alone would produce a startle response with exactly the same latency as that which was evoked by a sound loud enough to evoke startle in the first place. As we have already mentioned, these authors distinguished this type of response from a conditioned startle response on the basis of its latency and the variation of this latency as opposed to the constancy of the unconditioned response latency. They also noted that

while the unconditioned response was evoked either immediately or after few trials, many trials were necessary before the conditioned response appeared. These observations seem especially relevant to determining the true status of GSR to tone in any conditioning experiment.

In order to determine whether galvanic skin responses could be conditioned, using the latency of response as the chief criterion distinguishing conditioned responses from others, we carried out the following experiment.

METHOD

The *S*s were 19 student nurses from the Barnes Hospital School of Nursing. A uniform introduction to the experiment was given to the *S*s, put in general terms, and not telling them that they would receive shock. We made a point of asking *S*s not to let their fellow students know what went on in the experiment. The apparatus consisted of a Fels Dermohmeter connected to an Esterline-Angus chronograph, a series of three electronic timers, an Eico audiogenerator, and a shocking device. The *S* sat in a comfortable chair in a room separate from the experimenter. Recording electrodes in the form of zinc cups filled with an electrolyte jelly were placed on the *S*'s hand, one on the palm and one on the dorsum. Shock electrodes were placed on the *S*'s ankle, and the headphones through which the conditioned stimulus (a tone of 500 cps frequency and about 65 db. intensity) was presented, over the *S*'s ears. After a 5-minute rest period, the *S* was presented with the conditioned stimulus alone for 20 trials (adaptation period), the stimulus lasting 6.5 seconds and the intertrial interval varying at random between 30 seconds and 90 seconds. Next, 20 paired trials (conditioning period) were given at similar intervals, in which the unconditioned stimulus, a painful shock, was presented for .5 second 1.0 second after the end of the tone.

RESULTS AND DISCUSSION

The responses, or deflections, which appeared on the records could be separated into four categories. These were (*a*) spontaneous deflections occurring between trials, (*b*) and (*c*) unconditioned responses to the tone and to the shock, and (*d*) conditioned responses to the tone, or "conditioned anticipatory responses." In Figure 1 are shown the mean values and the ranges of latencies for responses of Categories *b*, *c*, and *d*, from *S*s taken alternately from a randomly arranged list of the group. Only the responses from the conditioning period are shown, but the characteristics of the unconditioned response to tone were the same in both periods. In particular, the relative constancy of the latency of response in any one *S* which can be seen in this diagram is true of both periods.

In drawing up our criteria for defining a "conditioned anticipatory response," we made use of the constancy of the unconditioned response latency. These criteria were: (*i*) the response had a latency which fell outside the range of latency for the unconditioned response to tone as recorded in the initial adaptation period, (*ii*) the response occurred at a time before a response to shock was physiologically possible, i.e., before a point 1.0 second after the onset of the shock, and (*iii*) the height of the deflection was at least equal to one of the small divisions on the vertical scale of the record paper. In practice, Criterion *i* usually meant that the response came at least 4.0 sec. after the onset of the tone.

Using these criteria, two of the investigators (JS and MS) went over the records independently, counting the responses at each trial. The reliability coefficient of their judgment of the total number of conditioned responses given by each *S* was .975. Perfect agreement was present in their judgments of the response number for 13 out of the 19 *S*s (68%), and agreement within a margin of one response in 17 out of the 19 (90%).

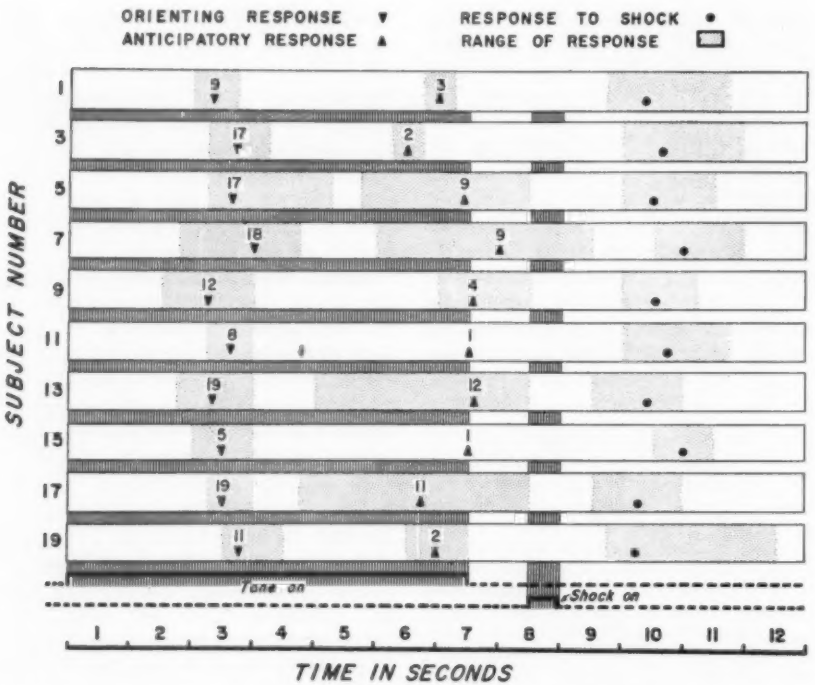
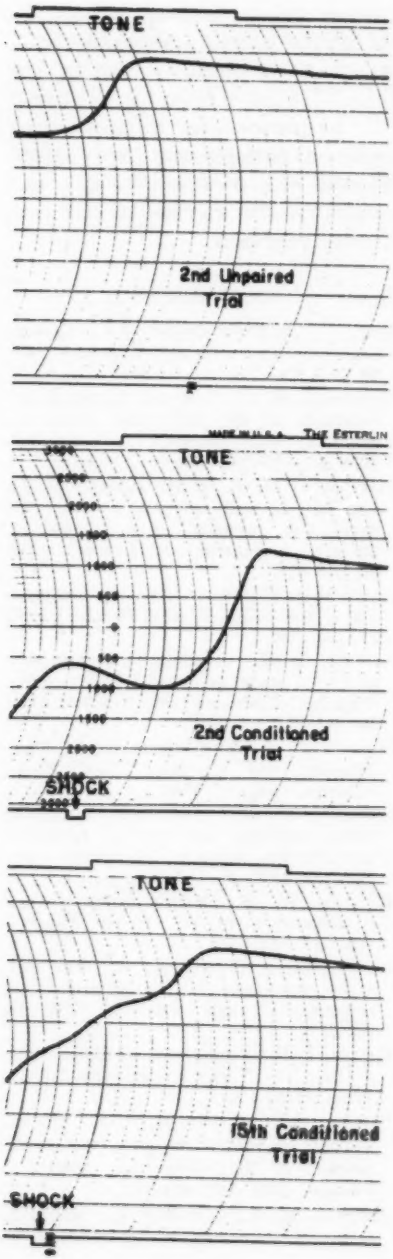


FIG. 1. The means and ranges of the latencies of the unconditioned responses to tone and to shock, and of the conditioned anticipatory responses, during the conditioning period. (Data are presented for alternate *S*s only. The numbers within the speckled areas refer to the number of responses on which the mean and range is based. Orienting response = unconditioned response to tone.)

Figure 2 shows representative samples of a record from one of the *S*s. The sample showing an example of the three responses occurring at one trial illustrates the fact that on occasion the conditioned response was part of a complex of deflections rather than being discrete. This has not actually made for any great difficulty in scoring the records, though it made it impossible to measure the amplitude of the deflections. It has, however, underlined the possibility that the deflection which we have called a conditioned response is part of the normal unconditioned response to tone or that it is a random spontaneous deflection. Responses corresponding to the conditioned responses of the second period did in fact occur in the initial adaptation period. Altogether there were 13 such responses, four *S*s making one each, two *S*s making two, and one *S* making five. These responses appeared to be randomly distributed over the adaptation trials, and there did not appear to be any correlation between the number of "conditioned responses" made in this period and the number made in the conditioning period. On the other hand, these responses were seen in the initial period in those *S*s



who had the highest rate of spontaneous deflections. The number of responses being small it was not feasible to treat these observed relationships statistically.

From these observations we tend to think that the occurrence of these "conditioned responses" in the initial period was the result of spontaneous GSR activity. Further evidence to support the idea that the responses occurring in the two periods were not of the same kind comes from the rates of spontaneous deflections in the two periods. We were able to count the spontaneous deflections in both periods in 16 out of the 19 *S*s, and found a total of 260 in the first period, and 395 in the second, giving a ratio 1 : 1.5. For the whole group the total of "conditioned responses" in the first period was 13, and in the second 77, giving a ratio of 1 : 5.9. From this comparison alone it would seem that the conditioned responses occurring in the conditioning period are not the product of spontaneous activity.

Figure 3 shows the number of responses given by the group at each trial of the two periods, leaving out unconditioned responses to shock. Gradual adaptation of the unconditioned response to tone is clearly present in the curve for the adaptation period, all of the 19 *S*s responding at the first trial, four at the last trial, and only one at the first trial of the conditioning period. At the second and following conditioning trials a marked recovery of this response takes place, and relatively little extinction is present at the end of the conditioning period.

The curve of the conditioned antic-

FIG. 2. Three samples from the record of a single subject. (The record reads from right to left, and each small division along the horizontal axis represents .5 second.)

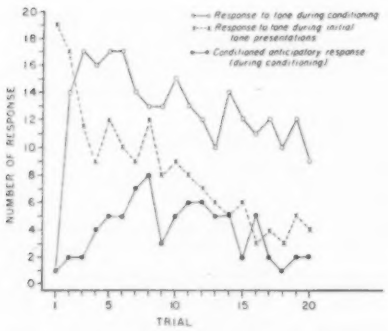


FIG. 3. The number of unconditioned responses to tone, and conditioned anticipatory responses, given by the whole group at each individual trial.

ipatory response approximates an inverted U shape, with a peak of eight *S*s responding at the eighth trial, and a decline to the low level of two responding at the twentieth trial. This shape is in obvious contrast to the usual learning curve, but it is also unlike the adaptation curves for unconditioned responses to tone. Our working assumption at present is that the "conditioned response" curve does in fact represent the acquisition of a true conditioned response, and that the extinction of the response is a function of massed reinforcement. A crucial experiment, then, for our hypothesis is a comparable experiment in which the intertrial intervals will be considerably longer, or in which aperiodic schedules of reinforcement are used.

In summary, we have put forward reasons for thinking that so far work on GSR conditioning has dealt with the adaptation and recovery of unconditioned responses rather than conditioning of responses. We have described criteria for defining true conditioned GSR in terms of the response latency, following methods used by previous investigators working with eyeblink and startle conditioning. The use of these criteria is illustrated in an experiment in which we tried to condition GSR in a group of 19 normal *S*s. Responses were obtained which could be distinguished from unconditioned responses and which had some of the characteristics of true conditioned responses.

REFERENCES

BERG, R. L., & BEEBE-CENTER, J. G. Cardiac startle in man. *J. exp. Psychol.*, 1941, **28**, 262–279.

BERNSTEIN, A. L. Temporal factors in the formation of the conditioned eyelid reactions in human subjects. *J. gen. Psychol.*, 1934, **10**, 173–196.

BITTERMAN, M. E., & HOLTZMAN, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615–623.

COOK, S., & HARRIS, R. Verbal conditioning of the GSR. *J. exp. Psychol.*, 1937, **21**, 202–210.

FRANKS, C. M. Conditioning and personality: A study of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1956, **52**, 143–150.

HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: Appleton Century, 1940.

HOWE, E. S. GSR conditioning in anxiety states, normals, and chronic functional schizophrenic subjects. *J. abnorm. soc. Psychol.*, 1958, **56**, 183–189.

HUNTER, W. S. Muscle potentials and conditioning in the rat. *J. exp. Psychol.*, 1937, **21**, 611–624.

LANDIS, G., & HUNT, W. A. The conscious correlates of the galvanic skin response. *J. exp. Psychol.*, 1935, **18**, 506–529.

PAVLOV, I. P. *Lectures on conditioned reflexes.* New York: International Univer. Press, 1928.

PROSSER, C. L., & HUNTER, W. S. The extinction of startle responses and spinal reflexes in the white rat. *Amer. J. Physiol.*, 1936, **117**, 609–618.

STEWART, M. A., WINOKUR, G., STERN, J., GUZE, S., PFEIFFER, E., & HORNUG, F.

Adaptation and conditioning of the GSR in psychiatric patients. *J. ment. Sci.*, 1959, **105**, 1102–1111.

THORPE, W. H. *Learning and instinct in animals.* Cambridge: Harvard Univer. Press, 1956. P. 72.

WHITE, C. T., & SCHLOSBERG, H. Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43**, 357–362.

WINOKUR, G., GUZE, S., STEWART, M., PFEIFFER, E., STERN, J., & HORNUG, F. Association of conditionability with degree of reactivity in psychiatric patients. *Science*, 1959, **129**, 1423–1424.

(Received November 17, 1959)

Psychological Review
1961, Vol. 68, No. 1, 68–80

EFFECTS OF SCORE TRANSFORMATIONS IN *Q* AND *R* FACTOR ANALYSIS TECHNIQUES[1]

DONALD M. BROVERMAN

Worcester State Hospital, Massachusetts

Controversy over the relationship of *Q* factor analysis (correlating people) to *R* factor analysis (correlating tests) has existed since Stephenson's introduction of *Q* technique in 1936. Stephenson (1936, 1952, 1953) has maintained that *Q* technique gives different results than *R* technique. Burt (1937), Eysenck (1953), and others, on the other hand, have argued that the two techniques are essentially transpositions of each other, i.e., factors of visual acuity and manual dexterity obtained from an *R* analysis will appear as factors of visually acute and manually dextrous individuals in a *Q* analysis of the same data. Cattell (1952a, 1952b), while agreeing with this latter position, has pointed out certain differences which make *R* preferable to *Q* technique.

This controversy reached its climax when Stephenson (1952) and Cattell (1952b) simultaneously published articles stating their respective positions. In retrospect, it seems that Cattell's position has been the more popular since over 95% of the factor analyses reported in the past seven years have employed the *R* technique.

The present paper presents new arguments and empirical evidence of important differences not previously recognized between the two techniques as they are commonly employed. The paper will try to show that the controversy over the *Q* and *R* techniques may be clarified by an examination of the effects of standardizing scores across the direction of correlation, i.e., standardizing columns when correlating rows, or vice versa. It will be argued that such procedures, which are customary in *Q* analysis, change the significance of scores, which, in turn, alters the inferences that may be drawn from the data.

[1] This paper was supported by the Dementia Praecox Research Project, Worcester State Hospital, and a research grant (M-896), from the National Institute of Mental Health, United States Public Health Service. The writer prepared this paper while a Public Health Postdoctoral Fellow in Clinical Psychology. Thanks are due to Melvin Feffer, Morton Wiener, and Jeffrey Caine for their invaluable comments during the preparation of this paper.

Standardizing Across the Direction of Correlation

Standardizing Columns When Correlating Rows

Cattell (1952a) has extensively discussed the effects of standardization of data on factor solutions. His general position is that standardization results in some loss of information but preserves the remaining data relationships. Thus, Cattell pointed out that if body measurements, e.g., weights, arm lengths, and finger lengths of individuals are compared, the "profiles" for all subjects (*S*s) will be similar since weight, measured in pounds, will be largest for everyone, while finger length, measured in inches, will be smallest. Accordingly,

> All people will correlate highly together in what we may call a "common species" factor . . . measuring various individual resemblance to common man. This factor is irrelevant for most differentiation within the species homo sapiens, and only after its extraction shall we come to the factors

of individual differences in which we are interested (Cattell, 1952a, p. 96).

Transformation of the above body measurements into standard scores, according to Cattell, eliminates the "species" factor, since weight and finger length now have distributions with similar means and ranges. Accordingly, the individual differences factor, which appeared second in the *Q* analysis of the raw data, should now appear first. The only effect of standardizing columns of measurements when correlating rows of people, according to Cattell, then, is the loss of the "species" factor.

Cattell's analysis of the effects of standardization on factor solutions is only partially correct. Besides losing the information pointed out by Cattell, standardizing columns of scores when correlating rows of people may bring out relationships between the rows which are not observable in the raw data. The reason for this is as follows:

When rows of raw data, such as body measurements, for two *S*s are correlated in a *Q* analysis, each *S*'s "profile" or distribution of scores is treated as a separate unit unconnected with the data of other *S*s. The correlation between the two *S*s, then, is based exclusively on the properties of each of these two distributions. However, when the columns of scores are standardized, each score in the person's profile is altered by a consideration, i.e., its relationship to the measurements of the other *S*s in that distribution, which did not enter into the above correlation. Thus, a score that was previously small in the raw data, such as finger size, may remain small in the standardized distribution if the finger is small among a population of fingers, or it may become a large score if it is a large finger in that population. These vagaries, which reflect the individual differences between *S*s, alter the relationships between rows.

Cattell's suggestion that this information about individual differences will be contained in the residuals of the raw score correlations after extraction of the "species" factor is incorrect, since this information, as indicated above, does not enter into the original correlations and the residuals contain only that part of the original correlations not accounted for by the first factor. Residual correlations cannot contain information not included in the original correlations.

The above argument is illustrated by the example suggested by Cattell, i.e., a table of body measurements. Table 1 presents imaginary raw data for 12 *S*s on 6 body measurements: height, arm length, foot size, neck size, and body weight. A *Q* analysis of these data,[2] as shown in Table 2, results in high positive intercorrelations of all *S*s, which, in turn, produce the massive "species" factor, predicted by Cattell. This factor is so massive that it includes over 99% of the total possible communality. The residual variance, which, according to Cattell, contains the individual differences information, would normally be considered too small to warrant further consideration. However, Table 2 also shows a minute second factor which was extracted from the remaining variance in order that comparisons might be made with the first *Q* factor

[2] Tryon's Cumulative Cluster Analysis (1958) was employed throughout this paper. This technique was used as a matter of convenience since it proceeds directly to an approximate simple structure factor solution similar to that obtained after rotation of axes in the more common centroid methods of analysis (Cattell, 1952a). However, while the factors discussed in this paper are all approximate simple structure solutions, the arguments apply to all factor solutions whether or not they are rotated to simple structure.

TABLE 1

BODY MEASUREMENTS, RAW DATA

	Height (feet)	Arm Length (inches)	Foot Size (inches)	Neck Size (inches)	Waist Size (inches)	Weight (pounds)
*S*1	6.3	36.0	12.0	17.0	36.5	195.0
*S*2	5.6	33.0	9.0	13.5	33.0	150.0
*S*3	5.9	35.0	11.0	14.5	32.0	155.0
*S*4	5.7	34.5	11.5	17.5	39.0	225.0
*S*5	5.1	31.5	7.5	15.0	34.5	180.0
*S*6	5.3	32.5	8.5	18.0	37.0	215.0
*S*7	5.4	32.0	10.0	16.5	38.0	210.0
*S*8	6.2	35.5	13.0	15.8	33.5	160.0
*S*9	5.8	34.0	9.5	14.0	34.0	165.0
*S*10	5.2	31.0	8.0	16.3	35.5	200.0
*S*11	5.5	33.5	10.5	16.0	36.0	205.0
*S*12	6.0	37.0	12.5	15.5	35.0	175.0

obtained after the columns of scores are standardized.

Table 3 presents the same body measurements with each column standardized to a mean of 50 and a sigma of 10. Table 4 shows the first *Q* factor obtained from this data, which, according to Cattell, is the equivalent of the small individual differences factor obtained from the raw data. However, unlike that diminutive factor, the individual differences factor obtained from the standardized scores is almost as massive (it contains 80%

TABLE 2

INTERCORRELATIONS OF ROWS AND *Q* FACTOR LOADINGS OF BODY MEASUREMENTS, RAW DATA

	Intercorrelations[a]												Factor Loadings	
	*S*1	*S*2	*S*3	*S*4	*S*5	*S*6	*S*7	*S*8	*S*9	*S*10	*S*11	*S*12	h_1	h_2
*S*1		9990	9992	9997	9998	9996	9996	9995	9996	9997	9999	9997	9995	0319
*S*2	−0002		9998	9977	9991	9977	9979	9996	9998	9981	9983	9996	9997	−0020
*S*3	−0003	0001		9980	9990	9978	9979	9999	9998	9981	9985	9999	9999	−0059
*S*4	0018	0004	−0004		9995	9999	9999	9984	9986	9999	9999	9987	9984	0558
*S*5	0011	0003	−0002	0020		9996	9996	9991	9996	9998	9997	9993	9992	0379
*S*6	0019	−0002	−0004	0033	0023		9999	9981	9985	9999	9998	9984	9982	0595
*S*7	0018	−0001	−0004	0032	0021	0034		9984	9987	9999	9999	9986	9983	0569
*S*8	0000	−0001	0000	0000	0000	0000	0000		9998	9989	9989	9999	9999	0001
*S*9	0002	0003	0000	0004	0006	0005	0006	−0001		9988	9990	9998	9998	0096
*S*10	0018	−0001	−0003	0031	0022	0034	0032	0000	0006		9999	9987	9984	0569
*S*11	0015	−0003	−0003	0027	0017	0028	0027	0000	0003	0026		9991	9989	0468
*S*12	0002	−0001	0000	0003	0002	0003	0003	0000	0000	0003	0003		9999	0053
	Total Communality of h_1 (h_1^2) = 11.980 Total Possible Communality = 12.000													

[a] Decimals omitted. Values above the diagonal are the original correlations. Values below the diagonal are the residual correlations after extraction of the first factor.

TABLE 3

BODY MEASUREMENTS
COLUMNS STANDARDIZED TO A MEAN OF 50 AND A SIGMA OF 10

	Height	Arm Length	Foot Size	Neck Size	Waist	Weight
*S*1	67	62	58	58	55	51
*S*2	49	47	45	33	38	33
*S*3	55	55	53	42	33	38
*S*4	51	53	55	62	67	67
*S*5	33	38	33	45	47	49
*S*6	42	45	42	67	58	62
*S*7	45	42	49	55	62	58
*S*8	62	58	67	49	42	42
*S*9	53	51	47	38	45	45
*S*10	38	33	38	53	51	53
*S*11	47	49	51	51	53	55
*S*12	58	67	62	47	49	47

of the total possible communality) as the first "species" factor obtained from the raw data. Examination of the factor obtained from the standardized scores reveals that it is defined equally by positively and negatively loaded *S*s. If one ignores the absolute size of the second factor extracted from the raw data and simply examines the relative weights of its loadings, it is seen to be defined predominantly by positively loaded *S*s. The average of the two negative loadings is only 6% of the average of the three largest positive loadings. Thus, standardization of scores not only brings out individual differences with much greater strength, but also presents a different picture of the relationships between *S*s than does a *Q* analysis from raw data. Clearly the *Q* factor obtained after standardizing columns of scores is a different factor than that obtained or obtainable from a *Q* analysis of the raw data.

Standardizing Rows When Correlating Columns

The above argument implies that standardizing rows, i.e., standardizing each individual's distribution of scores, will alter the column correlations employed in an *R* analysis. This is illustrated below:

An *R* analysis of the raw data in Table 1 produces the table of inter-

TABLE 4

INTERCORRELATIONS OF ROWS AND *Q* FACTOR LOADINGS OF STANDARDIZED COLUMNS OF BODY MEASUREMENTS

	Intercorrelations[a]												Factor Loadings
	*S*1	*S*2	*S*3	*S*4	*S*5	*S*6	*S*7	*S*8	*S*9	*S*10	*S*11	*S*12	h_1
*S*1		82	80	−90	−80	−68	−81	72	63	−74	−99	63	−848
*S*2	08		85	−90	−90	−99	−83	83	88	−95	−82	89	−872
*S*3	−04	−02		−98	−92	−83	−98	93	66	−92	−83	87	−996
*S*4	−07	−04	00		94	87	96	−92	−70	92	92	−86	982
*S*5	03	−11	−01	05		93	83	−98	−66	88	82	−82	909
*S*6	03	−26	00	05	17		78	−86	−87	94	68	−89	838
*S*7	02	03	00	00	−06	−04		−83	−69	92	85	−88	982
*S*8	−05	04	03	−03	−16	−10	06		56	−85	−75	82	−905
*S*9	04	28	−03	−02	−03	−29	−01	−07		−81	−59	74	−693
*S*10	05	−14	01	00	03	16	00	−01	−16		76	−98	932
*S*11	−25	−05	04	06	02	−06	−01	04	02	−06		−66	878
*S*12	−12	12	−01	01	−02	−15	−01	02	13	−16	11		−882

[a] Decimals omitted. Values above the diagonal are the original correlations. Values below the diagonal are the residual correlations after extraction of the first factor.

TABLE 5

COLUMN INTERCORRELATIONS AND *R* FACTOR LOADINGS OF RAW DATA

	Intercorrelations[a]						Factor Loadings	
	Height	Arm Length	Foot Size	Neck Size	Waist	Weight	h_1	h_2
Height		92	89	−05	−20	−34	953	−085
Arm Length	01		90	−03	−17	−30	960	−005
Foot Size	00	00		14	02	−12	939	139
Neck Size	−07	−05	12		83	88	021	910
Waist	−08	−05	14	83		95	−123	931
Weight	−09	−04	13	89	92		−267	974

[a] Decimals omitted. Values above the diagonal are the original correlations. Values below the diagonal are the residual correlations after extraction of the first factor.

correlations and the two *R* factors shown in Table 5. The first of these factors is defined by height, arm length, and foot size; the other by weight, waist size, and neck size. An *R* analysis of the standardized columns presented in Table 3 would produce essentially the same factors, since standardization of columns alters the column intercorrelations only to the extent that the column distributions are skewed.

Standardizing rows, however, radically alters the *R* factor solution of this data, since this transformation cuts across the line of correlation. The standardized columns of data in Table 3 are shown restandardized along each row in Table 6.[3] An *R* analysis of this matrix produces a single factor, shown in Table 7, which is defined positively by height, arm length, and foot size; and negatively by weight, waist, and neck size. Thus, whereas the *R* analysis of raw scores produced two independent factors, standardization across rows enables these data to be characterized by a single *R* factor. Once more, then, we see that standardization across the line of correlation changes the factor solution.

TABLE 6

BODY MEASUREMENTS STANDARDIZED FIRST BY COLUMNS, THEN BY ROWS

	Height	Arm Length	Foot Size	Neck Size	Waist	Weight	Means of Rows
*S*1	64	57	50	50	43	36	50.0
*S*2	64	57	52	39.5	48	39.5	50.0
*S*3	60.5	60.5	52	48	36	43	50.0
*S*4	36	43	48	52	60.5	60.5	50.0
*S*5	39.5	48	39.5	52	57	64	50.0
*S*6	39.5	48	39.5	64	52	57	50.0
*S*7	43	36	48	52	64	57	50.0
*S*8	57	52	64	48	39.5	39.5	50.0
*S*9	64	57	52	36	45.5	45.5	50.0
*S*10	45.5	36	45.5	60.5	52	60.5	50.0
*S*11	36	43	50	50	57	64	50.0
*S*12	52	64	57	39.5	48	39.5	50.0
Means of Columns	50.0	50.1	49.8	49.3	50.2	50.5	

DOUBLE VERSUS SINGLE CENTERED MATRICES

It is generally agreed, since Burt's mathematical demonstration (1937), that a doubly centered matrix, i.e., a matrix in which the columns and rows all have the same mean, will produce transposable factors. As indicated by

[3] Standardization of the rows of raw data leads to a matrix which cannot be analyzed. The raw score profiles of *S*s are so similar that all rows would have virtually identical standardized distributions. This, in turn, creates columns of scores with little or no variation to correlate; e.g., since weight is the largest raw score for all *S*s, the entire weight column would have the same standard score.

TABLE 7

COLUMN INTERCORRELATIONS AND *R* FACTOR LOADINGS OF DOUBLY STANDARDIZED DATA

	Intercorrelations[a]						Factor Loadings
	Height	Arm Length	Foot Size	Neck Size	Waist	Weight	A_1
Height		72	56	−64	−81	−90	−945
Arm Length	−07		46	−67	−75	−81	−832
Foot Size	−07	−09		−62	−56	−70	−664
Neck Size	−09	−18	−23		35	61	584
Waist	01	−03	01	−16		79	865
Weight	−02	−04	−08	07	−01		930

[a] Decimals omitted. Values above the diagonal are the original correlations. Values below the diagonal are the residual correlations after extraction of the first factor.

the marginal means in Table 6, a matrix of data which has been standardized first by columns and then by rows approximates a double centered matrix. A *Q* analysis of these data, then, should produce a factor which is the transposition of the single *R* factor already obtained from this matrix. However, it is unnecessary to perform this analysis since the *Q* analysis of the standardized columns of Table 3 constitute, in effect, a *Q* analysis of a doubly centered matrix. This is because the correlational process treats the distributions to be correlated as though they had the same means and sigmas. Since the columns in Table 3 were already standardized, correlation of the rows completed the process required to establish a doubly centered matrix. That this is so is indicated by Figure 1, where height, arm length, and foot size are seen to occupy higher positions in the distributions of measurements of *S*s loading positively on the *Q* factor obtained from Table 3 than weight, waist, and neck size measurements; with the reverse true of the negatively loaded *S*s. This factor (height, arm length, and foot size, versus weight, waist, and neck size), obviously, is the same factor as the one obtained from the *R* analysis of the doubly standardized matrix (Table 7), but expressed in terms of people. Thus, while a *Q* analysis of standardized columns is equivalent to an *R* analysis of a doubly centered matrix, both differ from an *R* analysis of either raw data or standardized columns.

Thomson (1939) has argued that *Q* correlations of raw scores involving different units are statistically meaningless since such scores do not form a metric continuum. Since that time, it has been standard practice in *Q* analyses (Babington-Smith, 1950; Eysenck, 1953) to standardize columns, thus producing an equivalence

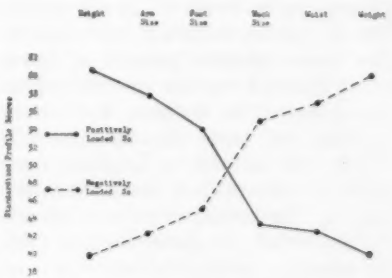


FIG. 1. Means of positively and negatively loaded subjects calculated from the standardized profile scores presented in Table 6.

of units, before correlating rows. As illustrated above, these procedures, in effect, create a doubly centered matrix. Hence, the usual *Q* analysis implicitly involves a doubly centered matrix, whereas the typical *R* analysis does not. In view of the differences in results obtained from single versus double centered matrices demonstrated above, it seems meaningful to conclude, then, that *R* and *Q* analysis techniques, as commonly performed, are different and give dissimilar results.

SIGNIFICANCE OF THE DIFFERENCE IN RESULTS

The above discussion suggests that the differences in results obtained from the different factor analysis techniques are a function of the different scores involved in single versus double centered matrices rather than the direction in which correlation proceeds. Evaluation of the differences in results, then, requires a careful consideration of these scores.

Normative versus Ipsative Scores

The usual *R* analysis of a single centered matrix involves "normative" scores (Cattell, 1944), or scores which reflect the standing of each *S* in each distribution with respect to other *S*s. These scores express interindividual differences in terms of group norms. Hence, factors obtained from normative scores describe patterns of interindividual relationships occurring within the group. For instance, a G factor implies that some *S*s are generally bright and do well on all tasks compared to other duller, poorer performing *S*s. Similarly, less general, specific factors refer to interindividual differences in abilities which vary independently of a G factor. *R* factors obtained from normative scores, then, reflect patterns of consistency within a *group* or population of *S*s. Hence, they might be appropriately termed "normative group" factors.

The usual *Q* analysis or *R* analysis of a doubly centered matrix, on the other hand, first standardizes columns of scores about a group norm, and then restandardizes each row of scores about a norm derived from the individual. Thus, these doubly standardized scores, termed "ipsative" by Cattell (1944), indicate whether the performance of an individual on a particular task is above or below his mean level of performance for the entire battery.[4] Ipsative scores, then, are scaled about intra-individual, rather than group, norms, and reflect intra- rather than interindividual variations in ability levels.

Factors derived from doubly standardized ipsative scores (in both *R* and *Q* techniques) also involve comparisons of *S*s. However, in this case, the comparisons are of the patterns of within-individual variation found in different *S*s. These factors, then, characterize the consistencies in *intra-person* organization which occur in the *S*s composing the sample. Hence, such factors might be appropriately termed "normative personological" factors.

It appears, then, that the two techniques produce factors which refer to different strata of behavior. Normative group factors describe consistencies in the composition of a group. Such characterizations may be useful

[4] It is also possible to develop ipsative scores from the responses of a single individual, e.g., galvanic skin responses under various conditions may be assessed relative to a base level characteristic of the individual. A matrix of such scores, however, would be as uncentered as a matrix of raw performance scores. Hence, double-standardization is necessary, regardless of whether behavior is assessed normatively or ipsatively initially, before similar results can be obtained in *R* and *Q* analysis techniques.

in developing a taxonomy of group behavior or in establishing the parameters along which relationships between performances of individuals vary. Normative personological factors, on the other hand, describe common patterns of intrapsychic variations in ability. Isolation of these factors is an appropriate goal for inquiries aimed at establishing the parameters of intrapsychic functioning.

In view of this, it is interesting to note that it is current practice in psychological research to make inferences concerning the organization of intrapsychic functioning from normative group factors (Burt, 1946; Thurstone, 1938). According to the arguments advanced in this paper, this practice is unwarranted and should lead to a considerably different theory of human functioning than would inferences drawn from normative personological factors. By the same token, experimental hypotheses may fail to receive empirical support if the data are not expressed in scores appropriate to the hypotheses.

An R *Analysis of Cognitive Data Expressed in Ipsative Scores*

The theoretical and empirical consequences of using different types of scores may be illustrated by the writer's experience with actual data from a battery of tests administered to 35 adult male members of a local church. These data were initially approached with hypotheses derived from the writer's previous work with two cognitive styles (Broverman, 1960a; 1960b). This work had indicated that individuals have stylistic tendencies to develop specialized abilities in either (*a*) novel or difficult conceptual behaviors, or (*b*) novel or difficult perceptual-motor behaviors ("Conceptual versus Perceptual-motor Dominance"). Accordingly, it was hypothesized that an *R* analysis of the test battery, which included several of each of the above mentioned classes of tasks, should produce a factor defined by inversely related difficult conceptual and perceptual-motor tasks.

The previous work had also shown that individuals vary, independently of the above style, in their ability to perform simple highly practiced tasks regardless of whether they are conceptual or perceptual-motor in nature ("Strong versus Weak Automatization"). Thus, a second factor defined by simple, automatized conceptual and perceptual-motor tasks was also predicted. However, the raw score *R* analysis of these data, reported by Podell and Phillips (1959), produced four independent *R* factors, none of which particularly resembled the predicted factors and so failed to support the above hypotheses.

Reconsidering the problem, it was noted that *S*s had been consistently used as their own controls in the previous work on cognitive styles. Hence, the differences observed between variously classified *S*s were actually based on intergroup comparisons of intra-individual variations in ability. This observation suggested that the factor hypotheses should be reformulated in intrapsychic or ipsative terms. Thus, it was hypothesized that the previously described factors would be obtained from an *R* analysis of ipsative test scores. Table 8 shows the factors obtained by this analysis. The first ipsative factor is defined positively by Thurstone's PMA Verbal Meaning Subtest, Nonsense Fluency Production, and Nonsense Fluency Order,[5] all of which may be reasonably termed novel or difficult conceptual behaviors; and negatively by a ratio

[5] Fuller descriptions of these tests may be found in Podell and Phillips (1959) and Brooks and Phillips (1958).

TABLE 8

TEST VARIABLES AND AUGMENTED FACTOR LOADINGS CALCULATED FROM IPSATIVE SCORES

Tests	h_1	h_2	h_3
Nonsense Fluency, Number Correct	994	−100	000
PMA, Verbal Meaning Subtest	981	−091	168
Nonsense Fluency, Order	972	−220	−078
Word Construction, Speed	748	226	−622
PMA, Reasoning Subtest	665	−661	−347
Stroop, Reading Color Names	635	637	435
Thurstone Gottschaldt	565	−650	−507
Goodenough Draw-a-Man	426	−642	−636
Porteus Mazes, Number Wrong Alleys	091	−953	−287
Subtractions	000	994	−102
Witkin Gottschaldt	000	−992	−119
Toothpick Design Construction, Number Produced Correctly	000	−635	−771
Card Sorting, Alphabetical Card Sorting	000	509	860
Stroop, Naming Color Hues	−078	950	−299
Stroop, Word-Color Interference	−203	922	328
Tapping Speed	−332	751	−220
Object Naming Speed	−402	915	000
Additions	−423	895	136
Card Sorting Ratio, Alphabetical/Nonalphabetical	−105	−164	−980
15 Clues, Number for Correct Solution	−379	−724	−575
Thurstone Two Hand Coordinator, Sum Both Hands Simultaneously	−382	−237	892
Wechsler Blocks	−302	−949	083
PMA, Space Subtest	−741	670	000
Toothpick Designs, % Correct	−924	−274	−264
Thurstone Two Hand Coordinator, Ratio Both Simultaneously/Sum Each Separately	−938	−319	133

Note.—Decimals omitted.

measure derived from Thurstone's Two Hand Coordinator, Accuracy of Toothpick Design Construction, and Thurstone's PMA Space Subtest. These latter tests are clearly novel or difficult perceptual-motor tasks. Thus, this factor corresponds closely to the first predicted factor.

The second ipsative factor is defined positively by speed of doing simple subtractions, speed of naming color hues, speed of reading the Stroop Word-Color Interference Card, speed of naming familiar objects, speed of doing simple additions, tapping speed, and speed of reading color names (see Fn. 5). Of these tests only the Word-Color Interference Test does not, on the face of it seem classifiable as a simple, automatized, conceptual or perceptual-motor task. However, the writer has previously interpreted performance on this card (Broverman, 1960b) as a function of the ability to make perceptual discriminations which should be a highly practiced, overlearned ability. The positively loaded tests defining this factor, then, support the hypothesis concerning automatized behaviors. Thus, while this analysis also delineated previously unsuspected classes of behavior, i.e., the negatively defining tests of the second factor and a small third factor, the results, in general, were in close agreement with the notions developed from the work with cognitive styles, whereas the results from the R analysis of raw

scores was not. Clearly, in this case, a successful outcome depended upon the use of appropriate scores.

Singly versus Doubly Defined Factors

The normative personological factors described above differ from the normative group factors of ability usually reported in the literature in another important aspect. Group factors of ability are typically defined only by positively loaded tests. The normative personological factors shown in Table 8, on the other hand, are defined with approximately equal strength by both negatively and positively loading tests. This difference in results will be discussed on two levels: (*a*) the numerical mechanics underlying the difference and (*b*) its implications for psychological theory.

The reason normative group ability factors are seldom defined both positively and negatively is as follows: a given group of *S*s will usually contain wide individual differences in general levels of ability. In this case, the individual mean levels of performance are distributed above and below the group mean. In such a sample, it is possible for the performances of a bright *S* on Tests A and B to be negatively related intrapsychically, and yet both performances may still be well above the group mean. Similarly, a dull *S* with the same inverse intrapsychic pattern of relationships between Tests A and B might give performances on both tests considerably below the group mean. Thus, the performances of the bright and dull *S*s contribute to a positive rather than a negative group relationship between these tests. Only if the mean level of performance of each *S* happens to correspond with the mean level of the group performance (in which case all *S*s would have the same general level of ability) would the inverse intra-individual relationships of Tests A and B occur as an inverse relationship for the group. Thus the presence of a G factor within a group acts to prevent the occurrence of both positively and negatively defined normative group factors.

In a previous paper (Broverman, 1960a) it was pointed out that ipsative scores may be used to control for inter-individual differences in G level. Presumably, a person's general level of performance on a large battery of tests will vary as a function of his general intellectual endowment. Hence, deviations of particular tasks above or below the person's general or mean level of performance may be considered ability variations independent of his G level. The use of ipsative scores in an analysis, then, may be considered a form of control for inter-individual variations in G level.

It should be noted that extraction of a general group factor from raw scores does not control G level in the same manner as ipsative scores. A G factor is based on correlations or covariances common to all columns. Extraction of this factor leaves the correlations specific to certain columns as residuals. However, these residual correlations are still based on normative scores and reflect normative group relationships which are independent of the G factor. As pointed out earlier, residuals cannot contain information not included in the original correlations. Since the status of a particular column score in an individual's profile distribution does not enter into the correlation between columns, this information can not be contained in a residual correlation. Ipsative scores, on the other hand, are based on precisely this information. Hence, these two methods of dealing with G are not comparable.

Control of differences in G through the use of ipsative scores, then, per-

mits inverse relationships between tests within the individuals composing the group to emerge as positively and negatively defined normative personological factors.

The implications of this difference in factor definition for theories of ability are considerable. Previous theorists have interpreted the occurrence of normative group factors defined exclusively by positively loading tests as an indication that intrapsychic organization consists of a number of independent dimensions on which *S*s differ only by being more or less proficient. Development, or an increase in the abilities of a given individual, then, simply implied increments in proficiency on the various dimensions of psychological functioning. Hence, development is treated in exclusively positive terms from this standpoint.

On the other hand, the occurrence of normative personological factors which are both positively and negatively defined suggest that certain abilities tend to be antagonistically related within the individual. An increase in ability in one class of behaviors apparently inhibits or lowers the level of ability in another behavioral class thus producing the inverse relationship. Of course, designation of one end of a doubly defined factor as positive or negative is arbitrary on the part of the factor analyst. Behaviors at a given end of a doubly defined factor represent dominant abilities for some *S*s and subordinate abilities for others. In other words, some *S*s will be stronger on the behaviors at one end of a factor, other *S*s stronger in the behaviors at the opposite end. Each doubly defined ability factor, then, may represent a "choice point" for the growing individual, i.e., at some point in his development a "decision" must be made on each bipolar factor as to whether he is to specialize in one or the other class of behaviors. Development viewed from this standpoint is a series of nodal points or forks in the road leading to specialization in one class of behaviors at the expense of another. Thus development is conceptualized in both positive and negative terms. It might be noted that this interpretation is similar to the notion of specialization in development presented by Werner (1957). Other investigators (Cattell & Baggaley, 1956; Sweney, 1959) have interpreted this specialization or directional aspect of ipsative scores as an index of motive strength.

Issues related to normative versus ipsative scores, then, include differences in results, differences in the type of inference possible from the results, and consequently, differences in the theoretical implications of the data. These differences, obviously, are of the greatest magnitude for research in psychology. Investigators employing factor analysis as a research tool should seriously consider whether the scores they employ are appropriate to their hypotheses, goals, or interests, if they are to avoid unwarranted conclusions, and possibly unnecessary failure.

SUMMARY

Although long a point of controversy, it is now generally maintained that factor solutions obtained from a *Q* analysis (correlating people) are essentially transpositions of factors produced by *R* analysis (correlating tests). The present paper takes issue with this position by presenting new arguments and empirical evidence of important differences between the two techniques, as they are commonly employed. The usual *Q* analysis is shown to involve, implicitly, a double centered matrix, i.e., a matrix in which all columns and all rows have the same

mean. In line with Burt's mathematical demonstration (1937) such a matrix is shown to produce equivalent factors by either technique. However, these solutions are shown to differ from the usual *R* analysis which, typically, employs a single centered matrix. As commonly performed, then, the two techniques give different factor solutions.

The above, however, suggests that the differences in factor solutions are a function of single versus double centered matrices rather than the direction of correlation. In turn, it is shown that the two matrices differ by employing different types of scores. Thus, the single-centered matrices customarily employed in *R* technique usually employ either raw data or standardized columns of test scores. In either case these scores, when correlated, are "normative" in that they reflect interindividual differences in terms of group norms. Similarly, relationships obtained from these scores are expressed with respect to group norms. Factors derived from normative scores, then, characterize consistencies within the performances of a group. Accordingly, it is proposed that such factors be termed "normative group" factors.

Double centered matrices, on the other hand, involve standardization of scores along rows as well as columns. Standardizing the rows of scores for each person produces "ipsative" scores, or scores which are scaled about intra-individual norms. Ipsative scores reflect intra-individual variations in ability. Accordingly, factors derived from ipsative scores represent patterns of intrapsychic ability organization common to the *S*s composing the group. It is proposed, therefore, that such factors be termed "normative personological" factors.

It is suggested that normative group factors are appropriate for inquiries aimed at establishing patterns of group behavior. Such information would be useful for problems involving group assessment, management, or prediction. On the other hand, isolation of normative personological factors seems the appropriate goal for inquiries aimed at establishing parameters of intrapsychic organization. From this standpoint, the usual procedures of making inferences concerning the organization of intrapsychic functioning, e.g., Thurstone's Primary Mental Abilities (1938), from normative group factors is unwarranted and should lead to different conceptions of human functioning than would inferences drawn from normative personological factors.

REFERENCES

BABINGTON-SMITH, B. Factor analysis: II. An evaluation of factor analysis from the point of view of a psychologist. *J. Roy. Statist. Soc., Series B,* 1950, **12**, 73–85.

BROOKS, MARJORIE O., & PHILLIPS, L. The cognitive significance of Rorschach developmental scores. *J. Pers.,* 1958, **26**, 268–290.

BROVERMAN, D. M. Cognitive style and intraindividual variation in abilities. *J. Pers.,* 1960, **28**, 240–256. (a)

BROVERMAN, D. M. Dimensions of cognitive style. *J. Pers.,* 1960, **28**, 167–185. (b)

BURT, C. L. Correlations between persons. *Brit. J. Psychol.,* 1937, **28**, 59–96.

BURT, C. L. *The factors of the mind.* London: London Univer. Press, 1946.

CATTELL, R. B. Psychological measurement: Ipsative, normative, and interactive. *Psychol. Rev.,* 1944, **51**, 292–303.

CATTELL, R. B. *Factor analysis.* New York: Harper, 1952. (a)

CATTELL, R. B. The three basic factor analytic research designs: Their intercorrelations and derivatives. *Psychol. Bull.,* 1952, **49**, 499–520. (b)

CATTELL, R. B., & BAGGALEY, A. R. The objective measurement of attitude motivation: Development and evaluation of principles and devices. *J. Pers.,* 1956, **24**, 401–423.

EYSENCK, H. J. *The structure of human personality.* London: Methuen, 1953.

PODELL, J. E., & PHILLIPS, L. A developmental analysis of cognition as observed in dimensions of Rorschach and objective test performance. *J. Pers.*, 1959, **27**, 439–463.

STEPHENSON, W. Introduction to inverted factor analysis with some applications to studies in orexis. *J. educ. Psychol.*, 1936, **27**, 353–367.

STEPHENSON, W. Some observations of *Q* technique. *Psychol. Bull.*, 1952, **49**, 483–498.

STEPHENSON, W. *The study of behavior.* Chicago: Chicago Univer. Press, 1953.

SWENEY, A. B. The objective measurement of motivation in children. *Amer. Psychologist*, 1959, **14**, 355. (Abstract)

THOMSON, G. H. *The factorial analysis of human ability.* London: London Univer. Press, 1939.

THURSTONE, L. L. *Primary mental abilities.* Chicago: Chicago Univer. Press, 1938.

TRYON, R. C. Cumulative communality cluster analysis. *Educ. psychol. Measmt.*, 1958, **18**, 3–35.

WERNER, H. The concept of development from a comparative and organismic point of view. In D. B. Harris (Ed.), *The concept of development.* Minneapolis: Univer. Minnesota Press, 1957.

(Received November 27, 1959)

THE BRITISH JOURNAL OF PSYCHOLOGY

Edited by BORIS SEMEONOFF

Vol. 51 Part 4 November, 1960 20s. net

T. H. PEAR. The accuracy of 'reproduced' conversation.

NORMAN F. DIXON. Apparent changes in the visual threshold: central or peripheral?

SATORU AIBA. The effects of stimulant and depressant drugs on the Bidwell phenomenon.

R. LYNN. Extraversion, reminiscence and satiation effects.

D. W. FORREST. Association between muscular tension and work output.

R. G. CHATTERJEA. The effect of distraction on time gap in the estimation of short duration.

HARRY M. B. HURWITZ. The effect of illumination conditions on the effectiveness of light-onset as a reinforcer: a test of the Bevan-Adamson reinforcement theory.

D. F. BROOK and A. W. HEIM. A preliminary note on the Brook reaction test.

CHRISTOPHER ALEXANDER. A result in visual aesthetics.

PUBLICATIONS RECENTLY RECEIVED.

The subscription price per volume, payable in advance, is 60s. net (post free).

Subscriptions may be sent to any bookseller or to the

CAMBRIDGE UNIVERSITY PRESS

Bentley House, Euston Road, London, N. W. 1

Elements of Psychology

by DAVID KRECH *and* RICHARD S. CRUTCHFIELD

University of California, Berkeley

"In its coverage of the major topics of psychology, there is depth as well as breadth. All material is lucidly organized and intelligently integrated. Sketches and diagrams are used freely, skillfully, and tastefully to illustrate empirical findings, details of experimental apparatus, and neurologic structure relevant to perception and emotion."—*Psychiatric Quarterly*. 758 pp. Illus. $7.00.

The STUDENT WORKBOOK by James L. McGaugh is a valuable supplement to *Elements of Psychology*. An Alternate Edition of the INSTRUCTOR'S MANUAL is also now available.

ALFRED A. KNOPF, *Publisher*

College Department 501 Madison Avenue New York 22